

doi:10.29285/actapinteriana.2019.5.61

Elme és tudatosság a fizikai világban: egy párbeszéd margójára

Schmal Dániel

Pázmány Péter Katolikus Egyetem
schmal.daniel@btk.ppke.hu

Schmal D. (2018): Elme és tudatosság a fizikai világban: egy párbeszéd margójára. Mind and Consciousness in the Physical World. Some Notes on the Margin of a Discussion. Acta Pintériana, 5: 61-89.

Abstract: The nature of the soul – or the mind as one prefers to call it nowadays – has long been one of the central issues of the trilateral discussions between sciences, philosophy and theology. During the discussions, the attention has often been focused on the metaphysical essence of the mind, because it is commonly agreed that not all kinds of philosophical and scientific positions can be reconciled with religious expectations. Accordingly, questions such as on what grounds and to what extent might the sciences lay claim on telling the final truth about the human nature, can legitimately be posed. If one is ready to accept the authority of the natural sciences, further issues ought to be clarified about human freedom, moral agency and the soul's independence of matter. Are these possibilities admitted into a physical universe governed by strict laws of nature and causality? These issues are often channeled into one central question about the nature of the mind: is the mind a material being or an immaterial, transcendent substance? In this paper I address and, to a certain extent, criticize this central question by drawing inspiration from the contemporary philosophy of mind. In a first step, I provide an overview of the most important principles of the field, the most well-known tenets and the difficulties inherent in each one. I maintain that in light of these scientific and philosophical insights the central question cannot be answered. The negativity of this result is just apparent. It sheds some light on the causes of the failure and thereby it opens the way towards another – by my lights more promising – approach. To make a step in this direction, in the second part of my essay I review a relatively new development in the contemporary philosophy of mind, the approach called 'situated cognition'. Finally, equipped with the results of these considerations, I make a proposal as to how the central question should be modified and reformulated.

A természettudomány, a filozófia és a teológia háromoldalú párbeszédének régtől fogva kitüntetett tárgyát képezi a lélek vagy – amint manapság mondani szokás – az elme természete. A figyelem a diszkussziók során legtöbbször az elme metafizikai lényegének problémája köré összpontosul, hiszen általánosan elfogadott meggyőződés szerint a lélekkel kapcsolatos vallási szempontok nem minden filozófiai állásponttal és nem minden természettudományos eredménnyel hozhatók összhangba. Ezért van kitüntetett jelentősége az olyan kérdések tisztázásának, mint hogy a természettudományok megismerési módszere milyen jogon és milyen mértékben tarthat igényt arra, hogy kimondja az emberrel és az emberi lélekkel kapcsolatos végső igazságot. S ha elfogadjuk a természettudományok tekintélyét, tisztázandó vajon a jelenségek okainak és törvényszerűségeinek tudományos feltárása hagy-e bármiféle teret a szabad akaratnak, az erkölcsi cselekvésnek, vagy a lélek testtől független létezésének. E kérdések gyakran abban az alapvető problémában összegződnek, hogy vajon az elme

anyagi létező-e, vagy immateriális, transzcendens valóság. Az alábbiakban ez utóbbi kérdés körüljárására – s mint látni fogjuk, kritikájára – vállalkozom a kortárs elmefilozófia (*philosophy of mind*) eredményei alapján. Első lépésben a teljesség igénye nélkül felvázolom a ma is folyó viták legfontosabb szempontjait, a legismertebb álláspontokat és ezek nehézségeit. Az áttekintés eredményeként azt fogjuk látni, hogy az imént feltett központi kérdés a mai természettudományos és filozófiai ismereteink fényében megválaszolhatatlan. Ez az eredmény azonban csak látszólag negatív. Számos olyan elvi kérdést tesz ugyanis láthatóvá, amely megvilágítja a kudarc okait, és utat nyit egy másfajta – reményeim szerint termékenyebb – megközelítés felé. Ezen az úton elindulva egy második lépésben a mai elmefilozófiák egy újabb áramlatát, a szituatív megismerés elméletét szeretném bemutatni, hogy azután ennek segítségével egy harmadik lépésben kísérletet tehessek a központi kérdés újrafogalmazására.

I. Tudat és fizikai világ a kortárs elmefilozófiában

1. A tudatosság kétféle megközelítése

Az elme létezésével kapcsolatban nem szorulunk tudományos bizonyítékokra, hiszen minden bizonyítékot megelőzően mindnyájunknak közvetlen hozzáférése van a saját tudatához. Mi több, felvethető, hogy az így kapott tudás nagyobb bizonyossággal áll rendelkezésünkre, mint bármilyen természettudományos ismeret. E körülményre René Descartes (1596-1650) hívta fel igen hangsúlyosan a figyelmet a 17. század első felében, amikor arra a belátásra jutott, hogy még ha mindaz hamisnak bizonyulna is, amit a tudományok állítanak a világról, e kudarc nem veszélyeztetné a saját létezésünk bizonyosságát (vö. DESCARTES 1994, pp. 33-34). Mert bármilyen kétellyel illetünk is mindent, amit a világról tudunk, saját tudatunk (az a tény tehát, hogy gondolunk valamit, még ha netán tévedünk is), valamint ezen keresztül a saját létezésünk (ha gondolok valamit, kell hogy létezzek) megrendíthetetlen igazságként áll előttünk. Descartes e tényből azt a következtetést vonta le, hogy minden egyéb ismeretünk is csak a saját megismerőrendszerünk közvetítésével adott számunkra, így a saját elmeállapotainkkal kapcsolatos bizonyosság foka minden esetben meghaladja annak a bizonyosságát, amit a világról csak ezen a megismerőrendszeren keresztül tudunk (vö. uo. pp. 42-43). A létezésemmel kapcsolatos ösbizonyosság (*cogito ergo sum*, „gondolkodom, tehát vagyok”) megelőz minden egyéb tudást, mert amíg a világ dolgaival kapcsolatos tudásom lehet hamis, az a tény, hogy valamit tudni vélek, cáfolhatatlan. Ezért – teszi hozzá Descartes – „nyilvánvaló módon fölismerem, hogy nincs semmi, amit könnyebben vagy evidensebben észlelnék, mint az elmémet” (uo. p. 43).

A lélek természetével kapcsolatos elképzelések e bizonyosság dacára véget nem érő vita tárgyát képezik a különböző kultúrákban. Hogyan magyarázható ez a különös körülmény? Hogyan lehetséges az, hogy a tudat egyszerre van a lehető legközelebb önmagunkhoz, és mégis oly kevésbé ismert? A válasz egy része abban a tényben rejlik, hogy a tudat e közvetlen megragadása speciális figyelmet igényel. Az esetek többségében nem a saját elménket, hanem a dolgokat figyeljük. Ahhoz, hogy elménk megismerhető legyen, ki kell lépnünk a dolgokhoz fűződő hétköznapi, vagy akár tudományos viszony keretei közül.¹ Míg a tudományokat művelve legtöbbször a körülöttünk lévő világ tárgyaira

¹ Edmund Husserl (1859-1938) a descartes-i (kartezianus) hagyományhoz kapcsolódva ezért az ún. „természetes beállítódás” feladásához kapcsolja a filozófiai gondolkodás kezdetét. A természetes beállítódás alapszerkezetéhez – amely meghatározza hétköznapi gondolkodásunkat – szorosan hozzátartozik a meggyőződés, hogy megismerésünk számára adottak az elmen kívüli világ tárgyai. A természettudományok művelése alapesetben (mindaddig, amíg a megfigyelő és a megfigyelt tárgy viszonya nem problematizálódik, mint a kvantumelmélet bizonyos interpretációiban) nem kívánja meg a szakítást e természetes beállítódással. Ez bizonyos értelemben arra a pszichológiára is vonatkozik, amely reflexív ugyan, de a megfigyelt

irányul a figyelmünk (például egy sejtsoportot, egy ásványt, egy fosszilis leletet vizsgálunk), addig saját tudatunk nem állítható önmagunk elé ilyen tárgyi formában. Egyszerűen nem tudunk kilépni magunkból, hogy mint egy izolálható tárgyat vegyük szemügyre önmagunkat. Amit tenni tudunk, az legfeljebb annyi, hogy *felfigyelünk* a tudatra mint olyasvalamire, ami a megismerés feltételeként mindig is jelen volt, amikor figyelmünk valami másra – egy fizikai jelenségre, egy sejtsoporra vagy egy fosszilis leletre – irányult. E tényt akkor tudjuk észrevenni s kellőképpen értékelni, ha megváltoztatjuk a figyelmünk irányát, és szemügyre vesszük, hogy mi is történik bennünk – hogyan is megy végbe az –, amikor egy tárgyat nézünk.

Ezt az irányváltást nevezzük reflexiónak. Mit is kell tennünk ahhoz, hogy ez a megismerésmód lehetséges legyen? Figyelmünk irányának megváltoztatása nem kívánja, sőt nem is engedi meg, hogy az eredeti tárgy – például a korábban szemlélt közet – kikerüljön a látóterünkől. Csupán annyit követel meg, hogy miközben továbbra is a közetet nézzük, figyelmünk egy részével konstatáljuk, hogy *ez a folyamat* zajlik bennünk: tudniillik a közetet nézzük. A megfigyelt tárgy továbbra is jelen van tehát, ami nem meglepő, hiszen a tudatot csak működés közben figyelhetjük meg, amint éppen valamire irányul, és nem állíthatjuk magunk elé a maga tiszta formájában mint tárgy nélküli tudatot. Ugyanakkor nem könnyű „nyakon csípni” ezen a módon a tudatot. A tudat ugyanis mint a reflexió tárgya nem egészen azon a módon van jelen, ahogyan korábban reflektálatlanul létezett. Eredetileg kizárólag a tárgyra koncentráltunk, s a közet tanulmányozásába merülve szinte meg is feledkeztünk magunkról.² Most, amikor a figyelmünket már nem a közetre, hanem a közetre irányuló megfigyelésre irányítjuk, a közet továbbra is jelen van ugyan, de immár nem a figyelem középpontjában álló tárgyként.

Ez a körülmény kérdéseket vet fel a reflexió megbízhatóságával kapcsolatban. Vajon amikor megfigyelői viszonyba kerülök a saját tudatállapotaimmal, e viszonyban valóban azt ismerem meg, ami – vagy amilyen – a tudat a megfigyelés előtt volt? Vajon maga a megfigyelés nem módosítja a megfigyelt dolog természetét? Hiszen a reflexió valamilyen módon mégiscsak tárgyiasít valamit, ami eredetileg – amíg figyelmem a közet szemlélésében merült el – nem tárgyi viszonyként volt adva számomra!³ S vajon honnan tudom, hogy amit így – önmagam felé fordulva – megfigyelek, azt helyesen figyeltem-e meg? Mihez tudom viszonyítani a megfigyelés eredményét? Vajon amikor másodszor figyelem meg ugyanazt önmagamban, helyesen emlékszem-e, hogy először mit is figyeltem meg? S vajon helyesen tudom-e kifejezni azt, amit megfigyeltem?⁴

Ezek és más hasonló nehézségek a 20. század elején arra az eredményre vezettek, hogy a behaviorista pszichológia elutasította a tudatállapotok szubjektív megfigyelését mint pszichológiai módszert (vö. WATSON 1970, pp. 39-58; PLÉH 2010, pp. 345-364). Ehelyett a lélektani problémák *természettudományos* módszerekkel történő vizsgálatához tért vissza, vagyis a lelki viselkedés tapasztalatilag megfigyelhető, kísérletileg manipulálható jeleit – az élőlényt ért ingereket, a viselkedéses választ és a kettő közötti korrelációkat – igyekezett laboratóriumi körülmények között feltérképezni. Eközben a belső, mentális szféra létezését vagy zárójelbe helyezte, vagy egyenesen elutasította.⁵ Az 1950-es évekre azonban fokozatosan kiderült, hogy az imént vázolt megközelítés nem

lélekállapotokat a világ tárgyaihoz hasonló – tárgyként minden további nélkül jelen lévő – megfigyelési objektumoknak tekinti (vö. HUSSERL 2000, pp. 45-46).

² Ez a megfélemezés ugyanakkor nem teljes: ha bárki megszakítja a vizsgálódásunkat, s rákérdez, mit csinálunk, azonnal készek vagyunk a válasszal, nem kell utólagos kutatásokat folytatnunk saját tevékenységünk felől.

³ „Elég gyorsan hátra tudok-e fordulni ahhoz, hogy elcsípjem önmagam?”

⁴ E nehézségekre különböző formában sokan felhívták a figyelmet a 20. században (lásd HUSSERL 2000, p. 46; WITTGENSTEIN 1998, § 298; GALLAGHER & ZAHAVI 2008, 3. fej.).

⁵ A módszertani korlátozás ugyanis gyakran kapcsolódott össze azzal a meggyőződéssel, hogy a viselkedés magyarázata után „nem marad vissza semmiféle megmagyarázhatatlan pszichikus reziduum, semmiféle tudatos

utasítható el minden további nélkül, s az elmével kapcsolatos gondolkodásban ismét polgárjogot nyert a belső elmeállapotokkal kapcsolatos igazságok elismerése. Mert bármit gondoljunk is az említett nehézségekről, annyi mindenképpen kiolvasható az iménti vázlatból is, hogy a mentális állapotok több alapvető ponton szembeállíthatók a fizikai jelenségekkel, amelyek a tudományos megfigyelés szokásos tárgyát képezik. Ezeket a tulajdonságokat vesszük szemügyre a továbbiakban.

2. A mentális állapotok megkülönböztető jegyei

Az úgynevezett *mentális állapotokat* vagy *elmeállapotokat* manapság három ismerv segítségével szoktuk leggyakrabban meghatározni, ezek az intencionalitás, a tudatosság és a privilegizált hozzáférés (vö. TÖZSÉR 2008, pp. 12-13). Mire utalnak e kifejezések?

a) Az *intencionalitás* szó annyit jelent, hogy elmeállapotaink valamilyen tárgyra vonatkoznak: a tudat valamiről való tudat. *Kívánni, látni, hallani, akarni* vagy *remélni* mindig *valamit* szoktunk, *félni, tartani, megijedni* csak *valamitől* lehet, *sóvárogni valamire* tudunk, és így tovább. Úgy tűnik, elmeállapotainktól elválaszthatatlan az a körülmény, hogy valamiről szólnak, valamire irányulnak.⁶

b) A *tudatosság* mibenlétét mindnyájan ismerjük, de nem könnyű definiálni, ráadásul a kifejezés többértelmű. Tisztázására az alábbiakban vissza fogunk térni, itt elegendő csupán annyit megjegyezni, hogy a mentális állapotok ismertetőjegyeként értett tudatosságot *fenomenális* vagy *kvalitatív* tudatnak nevezzük. E jelzők arra utalnak, hogy a tudatos elmeállapotok alanya számára *valamilyen* az adott állapotban lenni. A szék számára nem valamilyen, amikor eltörik a lába. Az ember számára valamilyen.

c) Végül elmeállapotainkat az különbözteti meg a fizikai állapotoktól, hogy egyes szám első személyű nézőpontból adódnak, és – ettől bizonyára nem függetlenül – privilegizált hozzáféréssel rendelkezünk hozzájuk. Mit jelent ez? Míg egy fizikai tárgy mindig a nyilvános tér része, azaz a fizikai világ tényeihez elvben bárki hozzáférhet, addig a saját mentális állapotaimmal kapcsolatban olyan tekintéllyel rendelkezem, amelyre más nem tarthat számot. Én tudom, hogy mit érzek, mire gondolok az adott pillanatban stb. Olyan közvetlen ismeretségben állok ezekkel a tartalmakkal, amilyen mások számára nem lehetséges. Ez egyszersmind annyit is jelent, hogy a tudatállapotok mindig perspektívához kötöttek: *számomra* jelennek meg, azaz szubjektívek, mert abból a nézőpontból adódnak, ahol *én* helyezkedem el, s amelyet nem foglalhat el más (lásd CRANE 2001, 2. fejt.).

Vitatott, hogy ez a három ismerv hogyan viszonyul egymáshoz. Összefüggenek-e vagy függetlenek egymástól? Van-e közöttük hierarchia, s ha igen, melyik a legalapvetőbb? Feltehetőleg van közöttük

sajátosság” (LASHLEY in KARDOS 1970, p. 74). E kettősség alapján szokás módszertani és filozófiai behaviorizmusról beszélni (lásd pl. SEARLE 2008, pp. 53-54).

⁶ Az *intencionalitás* szó modern kori használata a 19. század végén Franz Brentano (1838-1917) filozófiájával vette kezdetét. Brentano e középkori eredetű terminust a pszichológiai állapotok megkülönböztető jegyeként elevenítette fel (ez az ún. Brentano-tézis). Azt állította, hogy minden egyes mentális állapothoz tartozik egy-egy meghatározott tárgy, amelyre „mentális inegzisztencia” jellemző, azaz a megfelelő tárgy *intencionális értelemben* jelen van az elmében. Mit jelent ez? Amikor például a Napra nézek, a Nap mint a vizuális tapasztalat tárgya jelen van az elmében – de természetesen nem a maga fizikai valójában, hanem olyan tárgyként, amelyre a látás *irányul*, azaz *intencionálisan*. (Jól látható, hogy a szó e technikai jelentése nem áll közvetlen kapcsolatban az *intenció* – szándék – hétköznapi fogalmával.) Brentanótól azután kissé eltérő értelemben vették át a kifejezést a különböző 20. századi filozófiai iskolák, elsősorban Edmund Husserl és az őt követő fenomenológiai hagyomány, valamint az angolszász filozófia. A jelen írásban az utóbbi használati módot tartom szem előtt. Az angol filozófiai nyelvben az intencionalitás szinonimájaként az „irányultság” (*directedness*) és a „valamire vonatkozás” (*aboutness*) kifejezések is meghonosodtak.

kapcsolat, hiszen amikor a tudatosságról szólva azt mondtuk: a szék számára nem valamilyen, ha eltörik a lába, ezzel éppen azt állítottuk, hogy a szék esetében nem létezik privát nézőpont, ahonnan nézve számára bármi *valamilyen* lehetne. Szubjektum híján a szék esetében nem értelmezhető a „számára” kifejezés.

További problémát jelent a tudatosság és az intencionalitás összefüggése. Vajon minden mentális állapot jellemezhető mindkét ismérvvel? Úgy tűnik, léteznek olyan elmeállapotok, amelyek nem *valamiről* szólnak, vagyis nem intencionálisak, de tudatosak. Ilyen például a tárgy nélküli szorongás, amelyről nehéz lenne megmondani, hogy mire irányul. És megfordítva: léteznek olyan mentális folyamatok, amelyek tudattalanok ugyan, de tárgyakra irányulnak. Az alvajárók például feltehetőleg nincsenek tudatában a környezetüknek, ennek ellenére képesek tájékozódni a térben, kikerülik a tárgyakat. Vagy ide sorolható a vaklátás esete (vö. BLAKE & SEKULER 2004, p. 137), amely a primér látókéreg sérülésekor jön létre: a páciens ilyenkor nincsen tudatában annak, hogy bármit is látna a látótér egy meghatározott részén, viselkedése azonban azt bizonyítja, hogy az előtte lévő tárgy vizuális feldolgozása valamilyen szinten itt is végbemegy.⁷

Szerencsére e vitatott kérdésekben nem kell állást foglalnunk. Elég annyit megállapítanunk, hogy a mentális állapotokat e három ismertetőjegy definiálja együtt, külön-külön vagy valamilyen kombinációban. Ráadásul e három ismertetőjegy oly módon rögzíti a mentális jelenségeket, hogy egyben szembe is állítja őket a fizikai tulajdonságokkal. Amint a szék példáján láttuk, egy élettelen tárgy esetében nem feltételezzük, hogy állapotai önmagukon kívül létező dolgokra irányulnának (intencionalitás), bárminek a tudatában lennének, vagy lenne *számukra* olyasvalami, mint egy adott állapotban létezni.

3. A tudatosság kitüntetett szerepe a vitákban

A gondolkodó gépek kérdése

Az elme ismertető jegyei közül manapság elsősorban a tudatosság (b) áll a figyelem középpontjában (lásd CHALMERS 2004, pp. 3-13; CHALMERS 2005, pp. 7-18). Ez annak köszönhető, hogy a másik két ismerv számos kutató szerint visszavezethető az anyag természettudományosan kezelhető tulajdonságaira. Ha arra gondolunk például, hogy a robotok ugyancsak észlelik a környezetüket, és intelligens viselkedést produkálnak a tárgyakkal kapcsolatban, akkor jó érvek szólhatnak amellett, hogy ne tagadjuk meg tőlük az intencionalitást. Elvégre az önvezető autó éppúgy „látja”, hogy a lámpa piros (hiszen megáll előtte), és éppúgy észleli az útjába kerülő akadályokat (hiszen kikerüli őket), mint bármelyik ember.⁸ Hasonlóképpen kezelhetőnek tűnik a privát hozzáférés problémája is, hiszen – dacára annak, amint az imént a székről mondtunk – annak a nézőpontnak egy primitív formája, ami az ember esetében az egyes szám első személyhez kapcsolható, előállítható bármely fizikai vagy biológiai rendszerben, amelyik képes önmagát a környezetétől megkülönböztetni.

Az első (a) és a harmadik (c) ismérvvel kapcsolatos optimizmust az a remény élteti, hogy az elme *naturalizálható*, azaz annak ellenére, hogy az iménti ismérvek az elmét első megközelítésben szembeállítják a természettel, a mélyebb elemzés megszüntetheti a feszültségeket, s így az elmét végül sikerül a természet integráns részeként megérteni. A neves amerikai elmefilozófus, John Searle szerint a mai elmefilozófiának egyetlen nagy kérdése van: *How do we fit in?* (vö. SEARLE 2007, p. 4; SEARLE

⁷ Természetesen kérdés, hogy ezt a fajta öntudatlan percepciót hajlandóak vagyunk-e mentális állapotnak tekinteni. Ha igen, nyilván a viszony intencionális jellegére fogunk hivatkozni, ám számos filozófus nem ismeri el, hogy tudatosság nélkül beszélhetünk valódi intencionalitásról.

⁸ Természetesen ez az „éppúgy” egy sor problémát rejt magában. Lásd erről alább Searle *kínai szoba* gondolat kísérletét.

2008, p. 109). Azaz: miként illeszkedünk mi emberek a természetbe? A mesterséges intelligencia (a továbbiakban: MI) kutatását – azon belül is az úgynevezett *erős programot* – az a meggyőződés vezérli, hogy a választ az intencionális állapotokkal rendelkező, gondolkodó MI-rendszerektől remélhetjük. De vajon állítható-e egy gépről, hogy gondolkodik? E kérdést az MI-kutatások egyik kezdeményezője, Alan Turing fogalmazta meg egy azóta híressé vált, 1950-es cikkében.⁹ Turing választ is kínált e kérdésre egy olyan teszt alakjában, amely gépek és emberi lények különböző feladatok megoldásában nyújtott teljesítményét hasonlítja össze. A teszt lényege a következő. Egy személy kérdéseket tesz fel egy komputernek és egy embernek anélkül, hogy bármelyiküket is látná. Mindkettőtől valamilyen kódolt formában kap válaszokat, például képernyőn keresztül, így pusztán a válaszok formája alapján nem tudja, hogy géptől vagy embertől érkeznek. Akkor mondjuk, hogy egy gép gondolkodik, ha átment a Turing-teszten, azaz ha a kérdező a válaszok alapján bizonyos keretek között nem tudja megállapítani, hogy géppel vagy emberi személlyel folytat-e beszélgetést.

A Turing-teszt relevanciája nem magától értetődő. Erre a körülményre hívta fel a figyelmet John Searle, aki szerint attól, hogy egy gép átmegy a Turing-teszten, még nem intelligens a szó valódi értelmében, hiszen nem tesz egyebet, mint komputációs műveleteket végez anélkül, hogy *értene* azt, amit produkál. Searle e tézisének a *kínai szoba* című gondolat kísérlettel támasztotta alá (lásd SEARLE 1980, pp. 417-424; SEARLE 2008). Eszerint valakit bezárnak egy szobába, amelyben semmi más nem található, mint papír és toll, valamint egy könyv, amelyben utasításokat talál. A szoba falán két rés van, ezeken keresztül tud kommunikálni a külvilággal oly módon, hogy a réseken papírszeletek csúsztathatók ki és be. Tegyük fel, hogy az illető egy szót sem ért kínaiul, de képes arra, hogy amikor egy kínai jel ábráját becsúsztatják neki az egyik oldalon, odamenjen a könyvhöz, kikeresse benne a papíron látható jelet, majd az ott talált utasításnak megfelelően ráfessen egy másik jelet egy darab papírra, és kiadja a másik résen. Ha a könyv ügyesen van összeállítva, a kint álló kínaiaknak az a benyomásuk lehet, hogy aki a szobában van, kiválóan ért kínaiul. Ám intuíciónk azt súgja – s ez Searle konklúziója is –, hogy az illető semmit sem ért abból, amit csinál. Hasonló a helyzet a számítógépekkel, amelyek pusztán arra képesek, hogy szimbólumokat manipuláljanak bizonyos előre meghatározott szintaktikai szabályok szerint. Ha átmennek is a Turing-teszten, tudásuk nem a sajátjuk, hiszen nem értik, amit létrehoznak. Belső állapotaik legjobb esetben is csak utánozzák az emberi gondolkodást, de nem utalnak semmire.

Searle e gondolat kísérlete dacára általánossá vált az a meggyőződés, hogy az igazi nehézséget a tudat naturalizálásának programjával szemben nem az intencionalitás és a privát hozzáférés, hanem a tudatosság képezi. Ha elfogadjuk is, hogy a Turing-teszten átment MI-rendszerek gondolkodnak – azaz megengedjük, hogy intencionális viszonyban legyenek a külvilág tárgyaival –, még ekkor is ott van a tudatosság problémája, az a kérdés, hogy vajon mindez miért nem „sötétben” zajlik az ilyen gépekben, s hogy van-e *számukra olyan*, mint egy adott komputációs állapotban lenni (lásd CHALMERS 2004, pp. 3-13). E megközelítés az 1970-es évektől kezdődően a figyelem középpontjába állította a tudatosság kérdését. Ahhoz azonban, hogy ennek okait pontosabban lássuk, alaposabban szemügyre kell vennünk az elmefilozófiai vizsgálódások talán legfontosabb 20. századi paradigmáját, a funkcionalizmust.

⁹ Turing szerint a „képesek-e a gépek gondolkodni?” kérdés megválaszolása mindenekelőtt a *gép* és a *gondolkodás* kifejezések olyan operacionalizált meghatározását követeli, amely a gépi teljesítményt összemérhetővé teszi az emberi gondolkodással (vö. TURING 1950, p. 433). E meghatározás eszköze az az „imitációs játék”, amelyben egy ember és egy gép válaszol a feltett kérdésekre.

A funkcionalizmus

A funkcionalizmus az a meghatározó elmefilozófiai álláspont, amely a behaviorizmus meghaladásának szándékával az 1950-es évek végétől kezdődően megújította mind a pszichológiai, mind pedig a filozófiai kutatásokat (lásd HEIL 2004, pp. 139-141). Mint láttuk, a behaviorista tagadja – vagy módszertani okokból zárójelezi – a belső mentális szférát; kizárólag a receptorokra ható ingerek és a kimeneti viselkedés összefüggésére korlátozza a vizsgálatait abban a meggyőződésben, hogy ami ezek között van, az tudományos szempontból értékelhetetlen. A funkcionalizmus a behaviorista által elutasított mentális állapotok egyfajta rehabilitálása. Történetileg nézve a behaviorizmus hiányosságai – a viselkedés alaposabb magyarázatának igényei – kényszerítették ki újra azt a feltevést, hogy bizonyos organizmusok belső modelleket alakítanak ki magukban a külvilágról, azaz *szimbolikus reprezentációkat* tárolnak, s ezek manipulálásával különböző komputációs műveleteket végeznek, akár csak egy számítógép. Ezek az organizmusok egyéb jellegzetességeik mellett információfelvevő és -feldolgozó, -tároló – röviden kognitív (vagy megismerő) – rendszereknek minősülnek.

A funkcionalizmus e belső modelleket elemzi. A funkcionalista szerint az egyes mentális állapotok annak az oksági szerepnek a segítségével definiálhatók, amelyet az adott állapot egy kognitív rendszerben betölt. Mi például egy fájdalom? Ahelyett, hogy hosszas filozófiai vagy lélektani magyarázatokba bonyolódna, a funkcionalisták pragmatikusan közelítik meg a kérdést. A fájdalom – bármi legyen is a metafizikai természete – így jellemezhető: egy a test számára ártalmas hatás azt a vágyat kelti az alanyban, hogy az ártó állapotot szüntesse meg, és ez – *caeteris paribus* – elkerülő viselkedést eredményez. A fájdalom tehát a rendszerben betöltött oksági szerepével határozható meg. Általánosítva, a funkcionalisták szerint a következő három kérdésre kell válaszolnunk ahhoz, hogy egy mentális állapotot azonosítsunk.

1. Mi okozza az adott állapotot?
2. Mit okoz az adott állapot a kognitív rendszeren belül?
3. Milyen kimeneti viselkedést okoz?

Ha e három kérdésre válaszolni tudunk, akkor mindent elmondtunk, ami az adott mentális állapot szempontjából releváns. Vegyük észre, hogy ezek az adatok nem terjednek ki arra a kérdésre, hogy a szóban forgó állapot *miből van*. A funkcionalista megközelítés fontos mozzanata az a felismerés, hogy e kérdés teljesen közömbös, hiszen *a mentális állapotok mibenlétét az a szerep határozza meg, amelyet betöltenek a rendszerben*.

Világítsuk meg ezt a porlasztó vagy karburátor példájával (vö. BLOCK in HEIL 2004, p. 188). Mitől karburátor valami? A karburátor a belső égésű motorok számára a megfelelő üzemanyag–levegő arány előállítását végzi. E meghatározás tipikus funkcionalista definíció. Pusztán azt a szerepet rögzíti, amit el kell látnia valaminek ahhoz, hogy az legyen, ami: karburátor. Arról azonban e meghatározás semmit nem mond – nem is kell mondania –, hogy a szóban forgó eszköz miből készült, milyen a szerkezete stb. Készíthetjük fémből, fából, műanyagból, vagy akármi másból, amíg képes arra, hogy előállítsa a megfelelő üzemanyag–levegő arányt, addig karburátornak minősül. És hasonlóképpen: különféle technikai eljárásokat alkalmazhatunk, a lényegi kérdés ismét csak az, hogy a készülék képes-e ellátni a kívánt funkciót. Amíg erre képes, addig az alkalmazott technológia érdektelen abból a szempontból, hogy az eszköz karburátor-e.

Természetesen a technológiai különbségek vagy a gyártáshoz használt nyersanyagok más szempontból – például a gazdaságosság, a hatékonyság, a tartósság szempontjából – fontosak lehetnek, de ezek a szempontok nem érintik azt a lényegi kérdést, hogy karburátorral van-e dolgunk. Ez utóbbit kizárólag a funkció határozza meg. Hogy ez mennyire nem magától értetődő, az világossá válik, ha a karburátort egy ún. „természeti fajta”, mondjuk az arany mibenlétével hasonlítjuk össze. Az

arany definíciója szerint a 79-es elemszámú nemesfém. Az arany természetét az atommagban található protonok számával, és nem egy rendszerben betöltött szerepével rögzítjük. Az az elem, amelynek az atommagja 79 protont tartalmaz, arany, *bárhogyan viselkedjen is* (bármilyen funkciót lásson is el) egyébként. A meglepő viselkedés – mindaddig, amíg az anyag elemszáma 79 – nem ahhoz a következtetéshez vezet, hogy nem arannyal van dolgunk, hanem az arannyal kapcsolatos tudásunk felülvizsgálatára készítet. A karburátor példája ennek éppen ellenkezőjét mutatja. Ha meglepően viselkedik, azaz nem képes ellátni a szerepét, hajlamosak vagyunk azt mondani, hogy ezek szerint nem karburátor, nem pedig a karburátorral kapcsolatos tudásunkat módosítjuk.

A funkcionalizmus lényegi állítása e példák fényében így fogalmazható meg: az elmeállapotok karburátor-típusú létezők. Mivoltukat egy kognitív rendszerben játszott szerepük, és nem a metafizikai természetük rögzíti. E megközelítés forradalmi belátások forrása lett az 1960-as évektől kezdve. Először is egységes keretet biztosított a mesterséges intelligencia és a különböző – állati és emberi – elmetípusok kutatása számára. Hiszen ha azt a kérdést vizsgáljuk, hogy egy szervezet miként dolgozza fel a külvilágból érkező információt, miként szelektálja, raktározza, miként hívja azt elő, azt a további kérdést pedig zárójelbe tesszük, hogy e funkciók milyen anyagban valósulnak meg, akkor az emberi kogníciót ugyanúgy bizonyos műveletek kapcsolatrendszereként érthetjük meg, mint egy számítógépprogramot. Ahogyan egy számítógép esetében különbséget teszünk szoftver és hardver között, úgy a kognitív rendszerek esetében is különbséget tehetünk az elme lényegét alkotó funkcionális szerveződés és a fizikai hordozó között. Az természetesen korántsem magától értetődő, hogy egy komputer ugyanúgy vesz fel információt a környezetéből, s éppugyanúgy tárolja vagy hívja azt elő, mint egy biológiai szerveződés, de a teljesítmények igenis összehasonlíthatóak egymással. Ez a meggyőződés jól illeszkedik a Turing-teszt szellemiségéhez, s ez alapozta meg annak a kutatási programnak a máig tartó sikerét, amelyet *kognitív tudománynak* vagy *megismeréstudománynak* nevezünk (lásd PLÉH 2010, 21. fej.). Ez az interdiszciplináris kutatási program az információfeldolgozással foglalkozik, ennek során egyszerre kamatoztatja a pszichológia, a biológia, a filozófia, a nyelvészet, a számítástechnika, valamint számos más humán-, természet- és társadalomtudomány eredményeit.

A kognitivizmus tehát a behaviorizmus viselkedés-központú megközelítésével szemben a mentális reprezentációk vizsgálatára helyezte a hangsúlyt. A viselkedés ugyanis a kognitivista meggyőződés szerint nem egyszerű kondicionálás, hanem annak a képnek a függvényében változik, amit az élőlény kialakít magában a környezetéről. Az élővilág egy része *megismerőrendszerként* írható le, mert az élőlények modelleket, belső reprezentációkat alakítanak ki a világról, s e reprezentációk kialakulásának, tárolásának, előhívásának módozatai önállóan vizsgálhatók. A reprezentációk olyan funkcionális egységek, amelyek a világhoz fűződő viszony egyes elemeit tárolják, s más elemekkel kapcsolatokba lépve okságilag befolyásolják az élőlény viselkedését. A reprezentációk közötti összefüggés logikailag leírható szintaktikai szabályoknak engedelmeskedik.

Úgy tűnik tehát, hogy a funkcionalista elvben képes számot adni az emberi megismerőképeség teljes rendszeréről, arról, hogy az érzékszerveken keresztül érkező információ miként kódolódik biológiai regiszterbe, miként jönnek létre a megismerőrendszerben szimbólumok, ezek miként manipulálhatók, miként hívhatók elő, miként hatnak a viselkedésre és így tovább. Itt kell azonban visszakanyarodnunk a tudatosság fogalmához. Egyetlen dologról nem tud ugyanis számot adni a funkcionalizmus: hogy az általa leírt folyamatokat miért kíséri tudatosság, és miért nem zajlik mindez a „sötétben”.

A fenomenális tudatosság

Ezen a ponton nem halasztható tovább, hogy pontosítsuk a tudatosság fogalmának azt a jelentését, amelyre kritikusai szerint nem képes magyarázatot adni a megismeréstudomány. Amikor ugyanis azt kérdezzük, vajon a funkcionális folyamatok miért nem „sötétben” zajlanak, mindazt ki kell vonnunk a tudatosság fogalmából, amiről a funkcionalizmus sikeresen számot tud adni. A funkcionalista számára nem okoz nehézséget elmagyarázni, hogy a tudatos állapotok milyen oksági folyamatokon keresztül képesek belsőleg kódolni (szimbolizálni) a valóság különböző elemeit, milyen oksági kapcsolatokban állnak egymással, milyen viselkedéseket okoznak, és így tovább. Sőt a funkcionalizmus arról is magyarázatot tud adni, hogy a tudatos mentális állapotok milyen megfigyelhető különbséget okoznak a viselkedésben a nem tudatos feldolgozási folyamatokkal összehasonlítva. A tudatos folyamatok a tudattalanokkal szemben szeriális (sorba rendezett), lassú és energiaigényes folyamatok, amelyek ugyanakkor lehetővé teszik a különböző információk összekapcsolását egy egységes, koherencia-érzékeny, és számos kognitív alrendszer számára hozzáférhető regiszterben (lásd BAARS 1988/1995).¹⁰ Ezzel szemben a tudattalan kognitív folyamatok párhuzamos, gyors, de moduláris jellegük miatt zárt rendszerekben folynak, és nem érzékenyek az ellentmondásokra. Ezen a módon tehát számot tudunk adni arról, hogy *mi a tudatosság funkciója* például a tanulás során. Magyarázattal tudunk szolgálni az éber állapot, az alvás és az eszméletlenség különbségéről. Meg tudjuk magyarázni, hogy a megismerőrendszer elemei miként reprezentálják a valóságot (szemben például a tudattalan tárolással). Amikor tehát azt mondjuk, hogy a funkcionalista számára nehézséget jelent a tudat magyarázata, nem ezekre a kérdésekre gondolunk.

Miután mindezeket a funkcionálisan kezelhető elemeket – elsősorban a tudatállapotok reprezentatív és oksági funkcióit – kivontuk a tudatosság fogalmából, ami visszamaradt, az nem más, mint a tudatosság tiszta minőségi jellege. Ennek lényege abban a korábban említett megfigyelésben foglalható össze, hogy számomra mint megismerő számára egy adott mentális állapotban lenni nemcsak annyit jelent, amit az imént a funkcionalista leírás elmagyarázott – hogy ti. tájékozódni tudok a környezetemben, hogy számolni tudok az előttem álló tárgyakkal, manipulálni tudom őket, ki tudom őket kerülni stb. –, számomra mindezek az okságilag leírható tevékenységeken túl *valamilyen* e feladatokat végrehajtani, s *valamilyen* ezekben az állapotokban lenni. A tudatosság mintegy az oksági szerepeken túli maradék. A tudatállapotok e sajátosságát Thomas Nagel egy 1974-es cikke a következőképpen definiálta:

„[A]z a tény, hogy egy organizmus egyáltalán rendelkezik tudatos tapasztalattal, alapvetően azt jelenti, hogy van olyan, hogy valamilyen az illető organizmusnak lenni. [...] [E]gy organizmusnak akkor és csakis akkor vannak tudatos mentális állapotai, ha az illető organizmusnak lenni valamilyen – és pedig az organizmus számára valamilyen.

(NAGEL 2004, p. 4; kiemelés az eredetiben)

¹⁰Baars ebben látja a tudat egységének biológiai szerepét. Úgy véli, hogy a tudat létrehozásával az agy mintegy közös, *globális munkafelületet* teremt (ez az ún. *global workspace*-elmélet), amelyhez számos alrendszer járulhat hozzá információval, s amelyből számos kognitív alrendszer juthat információhoz. Mindennek elsősorban a tanulási folyamatban van szerepe, hiszen ismeretlen feladatokkal találkozva lényeges, hogy a probléma különböző aspektusait mintegy mesterségesen összefüggésbe tudjuk hozni egymással egy közös információs felület segítségével. „De még ha nincs is egy valódi végrehajtó fórum, a feldolgozó egységek egy megosztott csoportja rászorulhat olyan központi segédeszközre, amelyen keresztül a specializált egységek kommunikálhatnak egymással. Ezt a központi információcserét nevezzük 'globális munkafelületnek', 'táblának', vagy 'hirdetőfalnak' [...]. A 'munkafelület' csupán egy olyan memória, amelyben különböző rendszerek hajthatnak végre feladatokat, a 'globális' jelző pedig arra utal, hogy azok a szimbólumok, amelyek ebben a memóriában vannak, egy sor különböző feldolgozóegység között vannak megosztva.” (BAARS, 1988/1995, p. 87)

Ahhoz, hogy a funkcionalizmust bíráló kritikákat helyesen értsük, a tudatosságnak erre a fogalmára van szükségünk, hiszen – mint láttuk – a tudatnak léteznek olyan fogalmai, amelyek jól kezelhetők a funkcionalista elképzeléseken belül.¹¹ A tudatosságnak ezt a formáját *fenomenális* tudatosságnak nevezzük. A *fenomenális* jelző itt csupán annyit jelent, hogy amikor valaminek tudatában vagyok, az nekem *valamilyen*. A látásról például elmondható, hogy a szem optikai receptoraiban *ilyen és ilyen* elektrokémiai változások kezdődnek az elektromágneses spektrum egy meghatározott tartományába eső inger hatására, majd ezt az idegrendszer *így és így* dolgozza fel bioelektromos ingerületté, azután ez *ilyen és ilyen* neurális pályákon jut el az elsődleges, másodlagos látókéregbe... stb. – ez a történet azonban semmit sem mond arról, hogy nekem *milyen átélnem* a folyamatot. Ez a „milyen nekem átélnem” szemmel láthatólag olyan többlet az oksági történésekhez képest, amellyel a fizikai jelenségek nem rendelkeznek. Amikor egy pusztán fizikai jelenséget magyarázok, a dolog lényege szempontjából irreleváns, hogy számomra milyennek tűnik. Ezért a hő fogalma például jól magyarázható termodinamikai fogalmakkal, s az a szubjektív kérdés, hogy nekem *milyen*, amikor melegen van, semmit sem ad hozzá a fizikai jelenség magyarázatához. De amikor a magyarázandó éppen egy *érzés*, akkor nehéz volna azzal érvelni, hogy ennek természetét tekintve irreleváns, hogy az alany miként éli át (lásd NAGEL 2004, pp. 8-9).

A látás imént felidézett fiziológiai magyarázata tehát nem meríti ki mindazt, ami a látás, ehhez ugyanis az is hozzátartozik, hogy az adott organizmus számára – akiben a megfelelő fizikai-biológiai folyamatok végbemennek – látni milyen. Thomas Nagel példájával élve: ha a tudomány egyszer abban a helyzetben lesz is, hogy *mindent* tudni fog a denevérek életéről – pontosan ismerni fogjuk az idegrendszerük működésének minden kis részletét, az ekholokációs észlelés valamennyi sajátosságát – , egy dolgot akkor sem fogunk tudni, hogy ti. *milyen denevéreknek lenni*. Márpedig – érvel Nagel – ez a tapasztalat (maga az a tény, hogy az élő rendszerek egy része tapasztalattal rendelkezik, azaz számukra az élettelen dolgokkal ellentétben a világban lenni *valamilyen*) világunk magyarázatot igénylő ténye. Olyan „üledék”, amely akkor marad vissza, ha minden okságilag magyarázható tényezőt lepároltunk a tapasztalatból. A *fenomenális* jelző tehát a tudatosság e kvalitatív elemére, a tudatos tapasztalatok milyenségére utal. Egy latin eredetű műszóval a mentális állapotok e sajátosságát *kválénak* (többes számban: *kvália*) is szoktuk nevezni.

A magyarázati rés

A fenomenális tudatosság problémája arra a szakadékra mutat rá, amely az elme objektív, fizikai megértése és a tudatosság szubjektív élménye között húzódik. Ennek kapcsán szoktunk *magyarázati résről* vagy *szakadékról* (*explanatory gap*) beszélni. Bármilyen káprázatos eredményeket ért is el az agykutatás az elmúlt száz évben, mégsem világos, miképpen tudnak elektrokémiai folyamatok tudatos élményeket létrehozni a szó imént definiált, kvalitatív értelmében. Ha egy idegtudós a lehető legaprólékosabb elemzést végzi az idegrendszeren, s különböző agyi képalkotási eljárásokat alkalmazva lokalizálja a tüzelési mintázatokat, az ingerületátvivő anyagok terjedését, akkor számot fog tudni adni a funkcionális működésről (hogy miként kódolódik a kívülről jövő hatás, hogyan befolyásolja a viselkedést stb.), ám mindezek alapján képtelen lesz megmondani, hogy *milyen lehet átélni* a nekik megfelelő tudatállapotot. És megfordítva, egyes szám első személyű nézőpontból elemezve a tudatos tapasztalat élményét, sohasem lesz fogalmunk arról, hogy milyen idegrendszeri

¹¹ A tudatosság különböző fajtáinak megkülönböztetéséhez kiváló bevezetést kínál BLOCK in CHALMERS 1995/2002, pp. 206-218 (amely egyszersmind a ma legelterjedtebbnek tekinthető terminológiát is megalapozta); valamint KRIEGL in ZELAZO; MOSCOVITCH & THOMSON 2007, pp. 35-66.

folyamatok felelnek meg ezeknek az agyban. Bármelyik irányból induljunk is el, egy szakadék egyik partjára jutunk.¹²

Mindez nem teszi lehetővé, hogy *valamiféle* összefüggést teremtsünk a két oldal között. Megfigyelhetjük például a fenomenális tudat jellegét, s közben fMRI-vel mérhetjük az agyi folyamatokat. Ennek segítségével megalkothatunk egy olyan „szótárat”, amely rendszeres összefüggést teremt a szubjektív tapasztalat és az objektív tények egy-egy előfordulása között. Ám ennek segítségével pusztán a szubjektív és objektív tények együttjárását – azaz korrelációit – tudjuk rögzíteni, miközben továbbra sem értjük, hogy bizonyos idegrendszeri folyamatoknak miért *kell* bizonyos tapasztalati élményekkel együttjárniuk. A magyarázat szempontjából a kettő között továbbra is szakadék húzódik.

A korreláció (bizonyos jelenségek rendszeres együttjárása) nem kauzalitás. Ezt a tényt használják ki a funkcionális legismertebb kritikái. Ezek rendre arra mutatnak rá, hogy az elme funkcionális magyarázata semmit nem mond arról a *milyenségről*, amely a szubjektív tapasztalat sajátja. Arra az elméleti lehetőségre építenek, hogy ugyanaz a funkcionális rendszer *különböző* kvalitatív tudatossággal párosulva vagy akár mindenféle kvalitatív tudat nélkül is elgondolható, így a funkció a kvalitatív jellegre nem ad magyarázatot. A három leghíresebb gondolkísérlet ezzel kapcsolatban a következő:

- a) Az *invertált spektrum problémája*. Képzeljünk el két, funkcionálisan azonos megismerőrendszert, két elmét, amelyekben a piros és a zöld színnel kapcsolatos fenomenális tudat – tehát csakis a színészleléshez tartozó szubjektív élmény – fel van cserélve. Mit jelent ez? A két színnek megfelelő fizikai és biológiai folyamatok mindkét esetben azonosak, csak épp az ezekhez kapcsolódó kvalék tekintetében mutatkozik eltérés. Az egyik alany olyannak fogja látni a pirosat, mint a másik alany a zöldet és megfordítva. Kiderülhet ez valaha? Aligha, hiszen az illető is ugyanúgy tanulta meg a szavakat, mint mi, ő is azt fogja *zöldnek* nevezni, ami alul található a közlekedési lámpán, és *pirosnak* azt, ami felül. Ő is akkor fog megállni, ha a felső világít, s elindulni, ha az alsó. Az illető kognitív rendszere tehát funkcionálisan azonos a mienkkel, csak kvalitatíve különbözik, ám ez semmilyen oksági szituációban nem derülhet ki. Lehetséges ez? A funkcionista semmit sem tud mondani, ami ezt kizárja, ez azonban annyit jelent, hogy létezik az elmeállapotoknak egy olyan lényeges eleme (a fenomenális tudatosság), amely teljesen kívül marad a funkcionista magyarázat körén.
- b) A Kína-agy (*China brain*) gondolkísérlete (lásd BLOCK 1978, pp. 261-325). Képzeljünk el, hogy Kína több mint egymilliárd lakóját walkie-talkie vagy mobiltelefon segítségével úgy kötjük össze, hogy a hívások mintázata megegyezzen egy idegsejtcsoport tüzelési mintázatával. Ha az utóbbihoz tartozik egy kvalék (pl. egy fájdalomérzet vagy a kék szín érzete), vajon kapcsolódik-e *Kína teljes lakosságához* is? Az intuíciónk azt súgja, hogy a dolog képtelenség: hogyan is érezhetne Kína teljes lakossága fájdalmat vagy láthatna kék színt? (Vigyázzunk, nem az egyes emberek, hiszen az agykéregnek sem az egyes sejtjei látnak!) Ám a funkcionális szerint a két rendszer funkcionálisan egyenértékű, vagyis azonos mentális állapotban van. A funkcionális tehát nem lehet helyes, mert ellentmond annak a természetes intuíciónak, hogy a két állapot nem ekvivalens egymással: a neuronok tüzelése fenomenális tudatossággal jár együtt, míg a mobiltelefon-hívásokat nem kíséri tudatosság.
- c) A *zombi-érv*. Gondoljunk el egy zombit, aki sejtről sejtre megegyezik egy emberi személlyel, azaz az illető funkcionális hasonmása. Ám ha a funkcionális helytálló, akkor ugyanez a

¹²És megfordítva, egyes szám első személyű nézőpontból elemezve a tudatos tapasztalat élményét, sohasem lesz fogalmunk arról, hogy milyen idegrendszeri folyamatok felelnek meg ezeknek az agyban. Bármelyik irányból induljunk is el, egy szakadék egyik partjára jutunk.

fiktív lény mentális hasonmás is kell hogy legyen, hiszen a két rendszer okságilag ekvivalens (emlékszünk, a mentális állapotokat az oksági szerepükkel definiálta a funkcionista). A zombit azonban az különbözteti meg az embertől, hogy nincsenek tudatállapotai. De vajon lehetséges-e ilyen? Lehetséges-e, hogy ugyanazt a funkcionális működést az egyik rendszerben kíséri fenomenális tudat, míg a másikban nem? Úgy tűnik, e *logikai* lehetőség nem zárható ki. Márpedig ha a funkcionista nem zárhatja ki egy ilyen „filozófiai zombi” létezését, akkor ismét sikerült kimutatnunk, hogy a funkcionizmus tarthatatlan, mert nem érzékeny a zombi és a tudatos elme különbségére. Nem ad tehát teljes leírást az elméről, mert nem tud elszámolni a kvalitatív tudatosság fogalmával (lásd CHALMERS 1996, 3. fej.).

E példák alapján jól látható, hogy az elmúlt fél évszázadban miért a fenomenális tudat problémája került az elmefilozófiai diszkussziók középpontjába. Ha a funkcionizmus arra tett kísérletet, hogy egy közös, naturalista keretben értse meg az emberi és az állati kogníció, valamint a MI különböző problémáit, az elme különböző ismérvei közül a fenomenális tudat áll ellen legvilágosabban e törekvéseknek.

4. Mi következik a magyarázati rés létezéséből a tudatra nézve?

Eliminativizmus, azonosságelmélet

A magyarázati rés problémája lehetetlenné teszi a tudatosság fizikai redukcióját. Számos gondolkodó ebből arra a következtetésre jutott, hogy a tudat *nem fizikai jelenség* – még akkor sem, ha mindig együtt jár dinamikus agyi folyamatokkal, s így bizonyos értelemben függeni látszik a testtől. Ezt az álláspontot *szubsztancia-dualizmusnak* nevezzük. E kifejezés arra utal, hogy az embert két külön típusú, s egymástól elvileg független létező (szubsztancia) alkotja: a test, amelyik anyagi természetű, és az elme, amely anyagtalan. Ami a testi és mentális jelenségek állandó együttjárását illeti, már az ókorban megfigyelték, hogy a fej bizonyos sérülései kognitív képességek kiesésével vagy zavaraival járnak. Ez azonban nem mond ellent a dualizmusnak, annak a meggyőződésnek, hogy a mentális nem redukálható a fizikaira, s a fizikai és a mentális tulajdonságok *kategorikusan* különböznek egymástól. A dualista szerint a kategorikális különbség az együttjárás dacára arra utal, hogy a fizikai és a mentális nem azonos egymással.

Ugyanakkor ez a következtetés nem az egyedül lehetséges kiút a magyarázati rés okozta problémából. Számos filozófus és természettudós együtt tud élni a magyarázati rés jelenségével, s a legkevésbé sem tekinti olyan ténynek, amely az elme anyagtalan természetét bizonyítaná. A dualizmus legnépszerűbb kortárs alternatívája e tudósok körében az úgynevezett *fizikalizmus*. A fizikalista szerint minden, ami létezik, anyagi vagy – pontosabban – fizikai természetű (e pontosításra azért van szükség, mert a korábbi materializmussal szemben a mai fizikalista számos olyan magyarázó elvet elismer, amely a szó szigorú értelmében nem anyagi jellegű, ilyen pl. az energia). E megközelítés szellemében a magyarázati rés problémájára az egyik lehetséges magyarázatot az *eliminativizmusnak* nevezett fizikalista álláspont szolgáltatja. Az eliminativista szerint a magyarázati rés szubjektív oldalán megfigyelt jelenségek illúziók. Pusztán abból a tényből fakadnak, hogy egy sajátos pozícióból férünk hozzá a mentális jelenségekhez, ezért azok a mentalista (pszichológiai) kifejezések, amelyek ehhez a nézőponthoz kötődnek, a tudomány haladásával lassacskán „kiküszöbölhetőek” lesznek (innen az *eliminativizmus* elnevezés) a világ egy pontosabb leírásának a szótárából. Gondoljunk például az emlékezet jelenségére. A mindennapi életben mindnyájan tudunk arról, mit jelent emlékezni, mára azonban a kognitív kutatások részletes képet adtak azokról az idegrendszeri folyamatokról, amelyek lehetővé teszik az információ tárolását és előhívását az élőlényekben. Ennek során világossá vált, hogy amit mi emlékezetnek hívunk, az bonyolult alrendszerek, tárolási és előhívási mechanizmusok

sokfélesége az agyban. Nem létezik tehát olyan, hogy „az” emlékezet. Modulok sokasága létezik, s ezeknek a moduláris folyamatoknak a tudományos leírása végül feleslegessé teszi az emlékezet hétköznapi fogalmát.

Amikor az eliminativisták azt állítják, hogy a mentális fogalmak feleslegesek, ezen nem azt értik, hogy a szóban forgó jelenségek leírhatók egy egzaktabb tudományos nyelven, hanem azt, hogy a kognitív folyamatok e pontosabb leírása *kiküszöböli* e fogalmakat, hiszen meg tudja mutatni, hogy milyen heterogén működések sokasága teszi lehetővé azt a viselkedést, amit a hétköznapi ember az *emlékezet* szóval igyekszik visszaadni. Olyan tehát, hogy emlékezet, nem létezik – ez a szó éppúgy nem szerepel majd a tudományok egy teljesebb leírásában, mint ahogyan a *flogisztion* fogalma is eltűnt a modern kémia megjelenésével, vagy ahogyan a *boszorkányság* fogalma sem szerepel a modern pszichológia szótárában. Ezek a fogalmak nem redukció tárgyaivá váltak, azaz nem azt sikerült megmutatni, hogy amit régen *flogisztion*nak neveztek, az *valójában nem más mint...*, hanem az derült ki, hogy olyasmi, amit Stahl és követői a 18. század első felében *flogisztion*nak hívtak, *egyáltalán nem létezik* a dolgok végső leltárában. Az eliminativista hasonló értelemben igyekszik kiküszöbölni a magyarázati rés szubjektív oldalán található tudati jelenségeket. E megközelítés ugyanakkor abba a nehézségbe ütközik, hogy míg az említett példák valamilyen elméleti konstrukció eliminálásáról szólnak, addig a tudattal kapcsolatos szubjektív tapasztalatunk – mint láttuk – nem elméleti konstrukció, hanem nehezen tagadható elemi adottság.

Ígéretesebb választ kínál a fizikalizmus egy másik formája, az *azonosságelmélet*, amely a mentális állapotokat azonosítja fizikai állapotokkal. Az azonosságelmélet legizgalmasabb formáit ma azok az elméletek jelentik, amelyek azonosítják ugyan a mentális és a fizikait, de mégsem próbálják az előbbiket az utóbbiakra visszavezetni. Ezek az elméletek tehát – mint említettük – együtt tudnak élni a magyarázati rés tényével. Ezt a lehetőséget a *nem redukzív azonosságelmélet* vagy *nem redukzív fizikalizmus* képviseli, amely manapság az egyik legnépszerűbb elmefilozófiai álláspont. A nem redukzív fizikalizmus tézise szerint minden egyes mentális állapot azonos a központi idegrendszer egy fizikai állapotával. Ami miatt e megközelítés el tud számolni a magyarázati rés problémájával, az épp az elmélet nem redukzív jellege. Ez annyit jelent, hogy az említett azonosságot a szóban forgó fizikalisták nem tekintik magyarázó jellegűnek, azaz nem azt állítják, hogy a mentális tulajdonságok levezethetők a fizikai tulajdonságokból. Ellenkezőleg, a nem redukzív fizikalista úgy véli, hogy a mentális jelleg az anyag önálló, más fizikai ismérvekre visszavezethetetlen tulajdonsága.

A redukció elutasítása manapság nem csupán a magyarázati rés, hanem a *többszörös megvalósíthatóság* (*multiple realizability*) problémája miatt is népszerű. Mit jelent a többszörös megvalósíthatóság? Láttuk, hogy a funkcionalisták szerint egy mentális állapot azonossága szempontjából irreleváns, hogy az állapot „miből van”, vagy „miben valósult meg”, csak az számít, hogy mi okozta, s mire szolgál. Ugyanaz a mentális állapot tehát többféle hordozóban is realizálódhat. Ahogyan a karburátor többféle anyagból készíthető, úgy például egy elemi következtetés is, amit gondolkodás közben végzünk, *ugyanaz a mentális folyamat*, akár szén alapú vegyületben, akár (megint csak például) szilikonchipekben valósul meg. De a biológiai sokféleséggel és a neurális plaszticitással kapcsolatos kutatási eredmények is azt sugallják, hogy egy adott mentális állapottípus, például a fájdalom, különböző idegi környezetekben realizálódhat. Lehetséges, hogy *ugyanazt* a funkciót, amelyet az egyik élőlényben egy adott fajta neurális szerveződés valósít meg, egy másik fajban egy ettől eltérő szerveződés realizálja. Sőt, bizonyos korlátok között akár ugyanabban az agyban is különböző kérgi területek vehetik át egy adott funkció kiszolgálását az eredeti sérülése után. E megfigyelések alapján szokás arról beszélni, hogy egy funkcionális állapot különböző hordozókban

is megvalósulhat, azaz *többszörösen megvalósítható*.¹³

E körülmény több szempontból is figyelemreméltó. Egyrészt lehetetlenné teszi a redukciót, hiszen ha azt látjuk, hogy ugyanaz a mentális állapot típus hol x , hol y , hol pedig z fizikai hordozóhoz kapcsolódik, akkor egy mentális állapot ismeretében semmit sem fogunk tudni a mögöttes fizikai hordozóról. A nem reduktív fizikalista elfogadja ezt az elvet. Ezért pusztán annyit köt ki, hogy *minden egyes mentális állapot példány azonos egy fizikai állapot példánnyal*, de hogy melyikkel, azt nem szabja meg. Másrészt továbbra is fizikalista elméletéről van szó: nincsenek önállóan lebegő mentális állapotok, amelyek függetlenek az anyagtól. Minden egyes mentális állapot azonos egy fizikai állapottal, de hogy melyikkel, és milyen típusúval, az nem határozható meg reduktív módon. E megközelítést *példányazonosság-elméletnek* is szokás nevezni.

Ráépülés

Láttuk, de érdemes újra hangsúlyozni, hogy a többszörös megvalósíthatóság elvének fizikai elfogadása jól illeszkedik a funkcionálizmushoz. A funkcionálizmus metafizikailag semleges elmélet, amely a mentális állapotok oksági szerepeken keresztül történő megragadására tesz kísérletet. Metafizikai semlegessége annyit jelent, hogy miután nem érdekli, hogy miből van az, amit mentális állapotként azonosít, a funkcionálizmus összhangba hozható bármely olyan metafizikával, amelyik nem zárja ki a funkcionális magyarázatokat. Egy funkcionista hihet a lélek anyagtalanságában, nem véletlen azonban, hogy az elmefilozófusok többsége manapság nem dualista, hanem a nem-reduktív fizikalizmus álláspontjára helyezkedik, hiszen ez az elmélet megengedi a többszörös megvalósíthatóságot. Ha a fenomenális tudatosság és a magyarázati rés problémáját zárójelezzük, a funkcionálizmus és a nem reduktív fizikalizmus jól illeszkedik egymáshoz. A fizikalista ugyanis állíthatja azt, hogy az oksági szerepekkel (funkciókkal) azonosított mentális állapotok minden esetben valamilyen anyagi hordozóban valósulnak meg, ám hogy konkrétan milyen anyagban, az esetről esetre változhat.

A *megvalósulás* szó magyarázatot igényel. Amikor a nem reduktív fizikalista azt állítja, hogy a mentális állapotok agyállapottal azonosak, az azonosság egy speciális – a szó hétköznapi értelménél gyengébb – jelentésére gondol. Ezért fogalmaz úgy előszeretettel, hogy a mentális állapotokat fizikai állapotok „valósítják meg”. Az azonosság gyengébb értelme természetesen nem jelent engedményt az immaterializmus irányába. Továbbra is teljes azonosságról van szó, a mentális állapotok mindenestül fizikaiak, és semmi egyéb. A megszorítás pusztán annak szól, hogy a leírás mentális és fizikai oldala hogyan viszonyul egymáshoz.

Hadd világítsam meg ezt egy sok szempontból nem szerencsés, de szemléletes hasonlattal, a monetáris érték és a fizikai hordozók viszonyával. Az *ezerforintnyi* érték alapesetben nem létezik önmagában, valamilyen fizikai hordozó nélkül. Ám az ezerforintnyi értéket *megvalósíthatja* egy ezerforintos bankjegy, két ötszáz forintos papírpénz, esetleg tíz darab százforintos érme, és így tovább. Az érték minden esetben azonos valamilyen fizikai anyaggal, de többszörösen megvalósítható. Úgy is fogalmazhatunk, hogy az ezerforintnyi érték többféle materiális hordozóra *épülhet rá*. Ezért ezt a relációt *ráépülési relációnak* vagy latin eredetű kifejezéssel *szupervenienencia-relációnak* nevezzük.¹⁴ A

¹³ A kifejezést Hilary Putnam honosította meg a szakirodalomban (lásd PUTNAM in CAPITAN & MERRILL, 1967, pp. 37-48).

¹⁴ A monetáris értékkel kapcsolatos hasonlat azért nem szerencsés, mert az érték és a fémdarabok viszonya konvencionális: társadalmi intézményektől függ. Az elmeállapotok és az idegrendszeri szerveződések viszonya nem konvencionális, ezért ebben az esetben *természetes ráépülésről* beszélhetünk. A mentális állapotok az agy – mint bázis – különböző szerveződéseiben valósulnak meg, ezeken „szuperveniálnak”.

ráépülési reláció az azonosság egy speciális formája, amely megengedi a többszörös megvalósíthatóságot.

5. Status quaestionis

Az elmondottak arra utalnak, hogy a redukció lehetetlenségének kimutatása önmagában nemhogy nem cáfolja a fizikalizmust, de egyenesen összhangban van a mai legnépszerűbb fizikalista elméletekkel. Attól, hogy a tudatállapotok nem vezethetők vissza a fizikai hordozók tulajdonságaira, nem biztos, hogy metafizikailag különböznek tőlük, vagy bármilyen többletet hordoznak velük szemben. Még az egyes szám első személyű tapasztalat privát jellege sem zárja ki, hogy ez egyszersmind egy fizikai rendszerben megvalósuló állapot is legyen. Ez utóbbi – tehát a szubjektív karakter – és az azzal együttjáró „van olyan számomra, mint” jelleg fakadhat egyszerűen abból a tényből, hogy a szóban forgó igazságokhoz nem egy rendszeren kívüli nézőpontból, hanem belülről, a rendszer részeként, a leírandó rendszer egy sajátos pontján elhelyezkedve van hozzáférésünk. Hasonlattal élve: a szem nem látja magát, ezért az, *ami lát*, kategoriálisan különbözőnek tűnhet mindattól, ami a látás tárgya, azaz *amit látunk*. A tükörben látott szem is az „amit látunk”, és nem az „ami lát” kategóriájába esik, ám ettől még a kettő azonos egymással, referenciáik – amire az egyik és amire a másik leírás utal – összeérhetnek a valóság egyazon pontján.¹⁵

Ez természetesen nem azt jelenti, hogy megoldottuk volna a magyarázati rés kérdését. Szó sincs róla, hiszen fogalmunk sincs, hogy a két referenciális útvonal – az objektív és a szubjektív leírás utalásszerkezete – miként találkozhat össze egyetlen valóságban. De vegyük észre, ez a tény nem sokat mond a lélek természetéről, már csak azért sem, mert elsősorban tudásunk hézagaira mutat rá. Márpedig abból, hogy valamit nem értünk, nem könnyű érvet kovácsolni arra nézve, hogy a dolog lehetetlen. Colin McGinn ezzel kapcsolatban az emberi megismerőképeség *kognitív zártságáról* (*cognitive closure*) beszél (vö. MCGINN 1989/2004, p. 782). Egy rendszer kognitíve zárt egy bizonyos összefüggéssel szemben, ha a megismerőképeség egyszerűen alkalmatlan a probléma kezelésére. Ebben az értelemben mondhatjuk, hogy a csigák kognitív zártságot mutatnak a kvantumfizikával, vagy a macskák a piacgazdaság elméletével kapcsolatban (KRIEGEL in ZELAZO; MOSCOVITCH & THOMPSON 2007, p. 37). McGinn szerint semmilyen elvi megfontolással nem zárható ki, hogy létezik olyan természeti probléma, amelynek megértésével szemben az emberi elme az evolúció jelenlegi fokán kognitív zártságot mutat. Ebből azonban éppúgy nem következik, hogy az elme nem fizikai természetű, mint ahogy a csigák képességeiből sem következik semmi a kvantumfizika természetére nézve. A magyarázati rés egyszerűen a megismerésünkről mond valamit, s ebből nem könnyű a dolgok létezésére vonatkozó (ún. metafizikai) konklúziókat levonni.

Összefoglalva: azok az érvek, amelyek a tudatállapotok fizikai redukciója ellen szólnak, legjobb esetben is csak egy olyan fizikalizmus ellen lehetnek hatékonyak, amelyik a mentális állapotokat levezethetőnek tartja a fizikai állapotokból. A mai fizikalista elméletek jelentős része nem ilyen. Ezek az ún. *nem reduktív fizikalizmus* alesetei. A nem reduktív fizikalista – amint az elnevezés is mutatja – meg van győződve arról, hogy az elmeállapotok fizikai jelenségek, ugyanakkor arról is meg van győződve, hogy a tudatosság nem vezethető vissza az anyagnak azokra a tulajdonságaira, amelyeket a fizika rögzít. A redukció elutasítása azonban esetükben nem a lélek önállósága melletti érveknek, hanem a többszörös megvalósíthatósággal kapcsolatos empirikus eredményeknek tett engedmény.

¹⁵Természetesen hasonlatról van szó, hiszen a szó szoros értelmében nem a szem lát, hanem az elme, így a látó és a látott szem nyilvánvaló azonossága nem bizonyítja a fizikalista állítást. Jól szemlélteti azonban, hogy létezhet olyan viszony, amelyhez a szemlélő pozíciójából fakadóan nem tudunk objektív módon hozzáférni.

Ami tehát a *status quaestionis* illeti, a mai elmefilozófiai diszkussziók sem kizárni, sem megerősíteni nem tudják, hogy a tudat több lenne, mint az anyag. Ugyanakkor a mai kutatásokban nagy súllyal jelenik meg az az elv, hogy ahol nem rendelkezünk megbízható tudással, ott *jogos elfogultsággal* tekinthetünk a fizikalista előfeltevésekre. Jogos elfogultságról akkor beszélünk egy álláspont mellett, ha egy vita során nem az adott álláspont védelmezőjének, hanem az ellentétes álláspont képviselőjének kell bizonyítania az igazát. Ezt tükrözi például a bevett joggyakorlat, amelyben a vád képviselőjének kell bizonyítania a vádlott bűnösségét, és amíg erre hitelt érdemlően sor nem kerül, addig a vádlott ártatlannak tekintendő. Hasonló módszertani megfontolások a tudományok és a filozófia területén is gyakran érvényesülnek. A mai elmefilozófusok jelentős számban azért tekintik a fizikalista elvek melletti elköteleződést jogos elfogultságnak, mert a tudomány eredményeit tekintve a legsikeresebb episztemikus vállalkozás az emberiség történetében. Azok a módszerek és elvek (minden jogos tudományfilozófiai kétely dacára), amelyekkel a természettudósok dolgoznak, sikeres predikciókat tesznek lehetővé, s az eredmények technikai alkalmazása néhány száz év alatt megváltoztatta a Föld arculatát. E sikeresség felől nézve alapértelmezett pozíciónak az a meggyőződés tekinthető, hogy a magyarázandó jelenségeket a természeti kauzalitás részének és természettudományos módszerekkel megközelíthetőnek tekintjük. Ez a test–elme problémára alkalmazva annyit jelent, hogy a tudat *minden releváns tekintetben* a természeti világ részének tekintendő.

A jogos elfogultsággal és a természettudományos módszerek sikerességével kapcsolatos érvek elfogadása természetesen nem kötelező. A 21. században a népeségrobbanás, a globalizáció és az ökoszisztémára leselkedő veszélyek árnyékában éppen az utóbbi szempont (a sikeresség) nem magától értetődő, azt azonban látnunk kell, hogy az ezekkel kapcsolatos vita egy egészen más síkra emeli a vizsgálódást. Ezen a szinten már nem az elme természetéről, hanem általános módszertani problémákról, vagy épp alapvető kulturális és világnézeti beállítódásokról folyik a diszkusszió.

II. Elme és környezet

A továbbiakban a kortárs gondolkodás egy másik csapásirányát fogjuk szemügyre venni, amely nem a fizikai világtól való különbözőség (azaz végső soron nem a magyarázati rés) problémái felől közelít az elme természetéhez, hanem a fizikai világhoz fűződő pozitív viszonyt tartja szem előtt. Az itt következő „testközeli” vizsgálódások nem metafizikai jellegűek: nem feltételezik, hogy a lélek bármilyen értelemben nem-fizikai, ugyanakkor ezzel ellentétesen sem köteleződnek el.¹⁶ Az alább ismertetendő kutatásoknak mégis fontos relevanciájuk van az elme megértésének a szempontjából. Tézisszerűen megfogalmazva: nem arra utalnak, hogy a lélek „több, mint anyag”, hanem hogy „*mint anyag több*”, abban az értelemben, hogy egy elmével rendelkező fizikai rendszer önmagában is gazdagabb perspektívákat hordoz, mint amelyeneket az *anyag* szóhoz szoktunk asszociálni a metafizikai vitákban.

1. Internalizmus – externalizmus

Internalizmus

A lélekkel kapcsolatos klasszikus filozófiai elméletek *internalisták*. Ez annyit jelent, hogy a lélek vagy az elme állapotait minden esetben internálisan azonosítják. Bővebben: ahhoz, hogy egy mentális állapotot azonosíthassunk (rá tudjunk mutatni, meg tudjuk határozni, hogy melyikről beszélünk), az

¹⁶E sorok írójának személyes meggyőződése, hogy e kérdés mai ismereteink fényében eldönthetetlen.

internalizmus szerint semmi egyebet nem kell figyelembe vennünk, mint magát a szóban forgó állapotot, úgy, ahogyan az az elmében vagy (pontatlanabban, de hétköznapiasan fogalmazva) a „fejünkben” van. A mentális állapotok tehát az internalizmus szerint intrinzikus – belső – tulajdonságaik alapján azonosíthatók.

E megközelítés mellett a legegyszerűbb érvet a hallucináció szolgáltatja. Hallucináció során olyan érzéki tapasztalatokra teszünk szert, amelyeknek semmi nem felel meg a valóságban, s mégis, jellegüket tekintve megegyeznek a normál tapasztalattal – ezért is tudnak megtéveszteni. Minthogy a hallucinatorikus tapasztalatoknak semmi nem felel meg az elmén kívüli világban, kézenfekvő a feltételezés, hogy ilyenkor csupán saját belső állapotainkkal állunk megismerő viszonyban. E feltevés egy sor további érveléssel is megerősíthető. A külvilággal kizárólag az érzékszerveinken keresztül, az idegrendszer közvetítésével érintkezünk, a dolgokat tehát nem önmagukban látjuk, hanem úgy, ahogyan az idegrendszer aktivitása közvetíti őket számunkra. Ha valamilyen ok nem szándékos módon hat az idegrendszerre (például nem az érzékszerveket, hanem közvetlenül az agykéreg bizonyos területeit ingerli), előfordulhat, hogy összefüggő hallucinációkat tekintünk igaznak.

E lehetőséget feszíti végletekig az a Hilary Putnamtól (1926-2016) származó gondolat kísérlet, amely *agyak a tartályban* címmel vonult be a filozófia történetébe (PUTNAM 1981, pp. 1-22). Képzeljük el, hogy egy tudós kioperálja és tápoldatban életben tartja egy páciens agyát, majd az idegvégződéseket egy szuperszámítógépre kapcsolva mesterségesen hoz létre bennük impulzusokat. Amennyiben kellő ügyességgel stimulálja a sejteket, lehetségessé válik, hogy *bármilyen* tetszőleges érzéki tapasztalatot létrehozzon az agyban. (Gondoljunk például el, hogy a számítógép úgy programozza a tápoldatban tartott agyat, hogy az pontosan azokat a tapasztalatokat élje át, amiket a jelen sorok olvasója él át itt és most a képernyő előtt vagy a könyv fölé hajolva.) Tekintve az agy rendkívül összetett idegi szerveződését, a szinaptikus kapcsolatok hihetetlen számát, e kísérlet technológiailag kivitelezhetetlen. De lehetetlensége nem elvi, hanem mennyiségi természetűnek tűnik. Szerencsére nem is kell megpróbálkoznunk egy ilyen beavatkozással, a következtetés pusztán a gondolat kísérletből is levonható: mentális állapotaink attól azok, *amik*, hogy jelen vannak az elmében vagy az agyban, s ott úgy működnek, ahogy. A mentális állapotok korábban említett ismérvei – a tudatosság és az intencionalitás – internálisan jellemzik az egyes mentális állapotokat.¹⁷

Ugyanez az összefüggés az internalizmus egyik központi tételének, a *közös tényező elvének* a segítségével is megfogalmazható. Ehhez azonban tisztáznunk kell egy filozófiai szakkifejezést – a *veridikus* szó – jelentését. Veridikus, vagyis valóság-hű észlelésről akkor beszélünk, amikor az érzékelés megfelel annak, ahogyan a dolgok ténylegesen vannak. Amikor például a jelen sorok olvasója helyesen látja e sorokat, veridikus tapasztalatban van része, mert azok valóban ott vannak a könyvben vagy a képernyőn – ez választja el e helyzetet a hallucinációtól, amikor egy tápoldatban stimulálva látjuk pontosan ugyanezt anélkül, hogy létezne akár a könyv, akár a képernyő. Az előbbi esetben tehát megfelel valami a tapasztalatnak (a világ olyan, amilyennek észlelem), míg az utóbbi esetben nem. A közös tényező elve azt mondja ki, hogy maga a tapasztalat (vagyis az, amit a szó szoros értelmében látok) ettől függetlenül *mindkét esetben azonos*. Az internalista szerint akár veridikus az észleletem, akár hallucinálok, mindenképpen ugyanazt látom (a nyomtatott sorokat), vagyis ami a „fejemben” van, az mindkét esetben ugyanaz. A különbség nem a mentális állapotban, hanem a világban van. A vizuális tapasztalat olyan „közös tényező”, amely attól függően minősül veridikus észlelésnek vagy hallucinációnak, hogy *de facto* egy könyv előtt ülök vagy egy tartályban tartott agyként létezem.

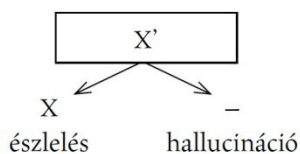
¹⁷ Az a gondolatom például, hogy „ma délután sütött a nap”, a tartalmánál fogva különbözik attól a gondolattól, hogy „a Föld a Nap körül kering”. Bármilyen kódolja is ezeket az intencionális tartalmakat az agyban, a megfelelő állapotok egy-egy belső tulajdonsága – vagyis egy *internális* tartozéka – teszi az egyes gondolatokat azonosíthatóvá, és megkülönböztethetővé egymástól.

Externalizmus

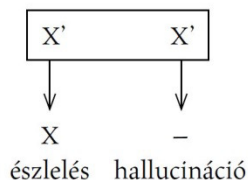
Ezzel a tézissel áll szemben az *externalizmus*, amely elveti a közös tényező elvét. Induljunk ki abból, hogy a valós észlelés és a hallucináció minőségileg valóban megkülönböztethetetlen egymástól. Ebből azonban nem következik az, hogy a szóban forgó két állapot azonos is egymással. Másként fogalmazva: az externalisták szerint nem következik, hogy a hallucinációnak és a valóság-hű észlelésnek lenne egy közös eleme, amit az egyik és a másik esetben is észlelek attól függetlenül, hogy megfelel-e neki valami a külvilágban, vagy sem. Az externalista szerint ilyen közös tényező nem létezik. A szóban forgó mentális állapot *vagy* veridikus észlelés, *vagy* hallucináció, de semmiképpen sem olyasmi, ami közös lehetne a két esetben. Ami miatt ez könnyen elkerülheti a figyelmünket, az az önmagunkra vonatkozó megismerőképeség hiányossága: mentális állapotaim viszonylatában nem feltétlenül vagyok abban a helyzetben, hogy különbséget tudjak tenni közöttük. Valahogy úgy, ahogy egy igazi és egy hamis Rembrandt-kép között sem feltétlenül tudok különbséget tenni – ettől azonban még nagyon is különböznek. Az externalisták szerint tehát függetlenül attól, hogy képes vagyok-e a különbséget tudatosítani, vagy az egyik állapotban vagyok, vagy a másik állapotban (azaz *vagy* észlelek, *vagy* hallucinálok), nem pedig egy észlelés/hallucinálás-semleges állapotban. Az ugyanis, hogy megfelel-e valami a tapasztalatomnak „ott kint, a világban”, nem közömbös abból a szempontból, hogy milyen állapotban vagyok „itt bent”.

Mindez azzal az eredménnyel jár, hogy egy adott mentális állapot természete nem ítéltető meg abból a szubjektív perspektívából, ahonnan számunkra adódik, azaz a természete nem merül ki az alany számára való adódásban.¹⁸ E tézis éles szakítást jelent azzal a korábban látott elvvel, amely szerint az alany, aki privilegizált hozzáféréssel rendelkezik saját mentális állapotaihoz, kitüntetett tekintélyre tarthat igényt mentális tartalmainak a megítélésében. Világítsuk meg az externalizmus álláspontját egy egyszerű példával. Képzeljünk magunk elé egy tízdolláros bankjegyet, és egy ezzel külsőleg megegyező, tökéletes hamisítványt. Ekkor a szóban forgó papírok mindegyike *vagy* tízdolláros lesz, *vagy* hamisítvány. Nem létezik olyan közös tényező, amely önmagában még nem dollár, de nem is hamisítvány. Még ha a kettő bármely szemlélő számára tökéletesen megkülönböztethetetlen is, biztos, hogy vagy az egyikről van szó, vagy a másiktól. A két álláspont közötti különbség az alábbi ábrán szemléltethető:

INTERNALIZMUS



EXTERNALIZMUS



¹⁸ Ezt hagyományosan úgy szoktuk megfogalmazni, hogy az externalista szerint egy mentális állapot esetében nem igaz az *esse est percipi* elv, azaz az állapot léte nem merül ki abban, hogy a tudatában vagyunk. Az *esse est percipi* George Berkeley (1685-1753) ír filozófus idealista alapelve, amely szerint „*létezni annyi, mint észlelve lenni*”. Míg Berkeley e tételt az anyagi világra is kiterjesztette, azaz a fizikai világot is a szellem észleleteivel azonosította, addig az externalizmus ugyanezt az elvet az ideák vonatkozásában is tagadja.

Az externalista tanítás azon a belátáson alapul, hogy mentális állapotaink jelentős részének – például az érzeteknek, észleleteknek – alapvető szerepe a *referálás*, vagyis az a tulajdonságuk, hogy utalnak valamire a világban (reprezentálnak valamit), s ezáltal összekötnek bennünket a valóság tárgyaival, eseményeivel. Amennyiben e szerepüknek nem tesznek eleget, nem valamilyen részleges hiányosságban szenvednek, hanem alapvető természetüknek nem felelnek meg – egyszerűen nem azok, amiknek látszanak. Észleletnek tűnnek ugyan, valójában azonban nem azok, hanem hallucinációk.

Az externalizmus melletti egyik legszemléletesebb érv – az úgynevezett *Ikerföld-érv* – ugyancsak Hilary Putnamtól származik (lásd PUTNAM 1975/2003, pp. 222. skk.). Képzeljük el, hogy valahol a világegyetemben létezik egy másik Föld (nevezzük Ikerföldnek), amely minden tekintetben pontosan megegyezik a Földünkkel, s így minden ugyanúgy játszódik le rajta, mint a Földön. Az Ikerföldön tehát minden tárgynak (növénynek, embernek stb.) megfelel egy ikerföldi hasonmása, aki vagy ami pontosan ugyanolyan és pontosan ugyanúgy cselekszik, mint a földi példány. (Miközben tehát az olvasó ezeket a sorokat olvassa, ikerföldi mása ugyanezt teszi egy ikerföldi könyvben.) Egyetlen kis különbség választja el csupán a Földet az Ikerföldtől, az, hogy az ikerföldi víz kémiai összetétele valamilyen rejtélyes oknál fogva nem H_2O , hanem bármi más, mondjuk XYZ.¹⁹ Minthogy azonban a felszíni tulajdonságok megegyeznek – mind a víz, mind az ikerföldi színtelen, szagtalan folyadék –, erről a mélyszerkezeti különbségről senki nem tud. (Putnam a gondolatkísérletet a 18. század elejére, a Lavoisier-féle kémiát megelőző korszakba helyezi.) Végezetül képzeljünk el a Földön egy Oszkár nevű egyént, aki egy pohár vízre mutat, és ezt a kijelentést teszi, vagy egyszerűen ezt gondolja: „*Íme, egy pohár víz.*” Ugyanebben a pillanatban az Ikerföldön is létezik egy másik, ugyancsak Oszkár nevű egyén, aki ugyancsak egy pohár vízre mutat, és szintén ezt gondolja: „*Íme egy pohár víz.*” Kérdés: vajon Oszkár és iker-Oszkár ugyanolyan mentális állapotban vannak-e?

Putnam szerint intuíciónk azt súgja – helyesen –, hogy nem. Oszkár kijelentése vagy gondolata ugyanis egy H_2O összetételű anyagra utal (referál), míg iker-Oszkáré egy XYZ összetételű anyagra. A fejükben ugyanaz van ugyan, hiszen a víz fogalmát mindketten a felszíni tulajdonságok alapján alakították ki („színtelen, szagtalan folyadék”), a két kijelentés mégis különböző. Azért nem lehetnek azonosak, mert nem ugyanarra az anyagra vonatkoznak – nem ugyanaz az anyagtípus elégíti ki (teszi igazzá) az egyiket és a másikat. Oszkár gondolata akkor igaz, ha H_2O van előtte a pohárban, de nem igaz, ha XYZ van benne, azon egyszerű oknál fogva, hogy Földünkön a víz – akár tudunk róla, akár nem – *de facto* H_2O , és nem XYZ. Ez egy olyan tény, amely független attól, hogy „mi van a fejünkben”. Mentális állapotaink tehát (jelen esetben a víz fogalma) olyan tényekre is érzékenyek, amelyek hozzájuk képest külsők.

A putnami gondolatkísérlet konklúziója tehát úgy szól, hogy ami egy gondolat vagy mentális állapot azonosságát adja, az nem ennek az állapotnak valamilyen internális tulajdonsága, hiszen ahhoz, hogy azonosítani tudjuk, a külvilág egy részét is tekintetbe kell vennünk. Oszkár és iker-Oszkár hiába rendelkezik minden belső tulajdonságát tekintve azonos fogalommal a vízzel, a két gondolat mégis típusosan különbözik egymástól. (Típusosan: tehát nemcsak az különbözteti meg őket, hogy az egyik Oszkár, a másik iker-Oszkár fejében van, hanem *más fajta* a kettő – az egyik „színtelen-szagtalan-folyadék-ami- H_2O -ra-referál”, a másik „színtelen-szagtalan-folyadék-ami-XYZ-re-referál”). Hasonló

¹⁹ Az XYZ megnevezés Putnamtól származik, s csupán egy tetszőleges kémiai összetételre utal. Putnam a gondolatkísérlet konklúzióját először szemantikai összefüggésben a „víz” szó jelentésével kapcsolatban fogalmazta meg: „*Ha valaha egy földi űrhajó meglátogatja az Ikerföldet, először azzal a feltevessel fognak élni, hogy a 'víz' ugyanazt jelenti a Földön és az Ikerföldön. Ezt a feltevést azonban korrigálják, miután felfedezik, hogy a 'víz' az Ikerföldön XYZ, s a földi űrhajó valami ilyesféle beszámolót készít: az Ikerföldön a 'víz' XYZ-t jelent.*” (PUTNAM 1975/2003)

A jelentések tehát „nem a fejben vannak”. Innen csupán egy lépés magukra a mentális állapotokra is kiterjeszteni a konklúziót.

következtetésre juthatunk a tízdolláros példánk alapján. Tíz dollárnak lenni nem a papír és a festékek intrinzikus tulajdonsága. Ami *ezt* a darab papírt tízdolláros bankjeggyé teszi, az nem önmagában a festék és a papír, hanem az, hogy a *Federal Reserve* bocsátotta ki (s nem magam állítottam elő a konyhában).

Az utóbbi példa egy további összefüggésre is felhívja a figyelmet. Ha a dollár (részben) attól az, ami, hogy oksági kapcsolatban áll a kibocsátásra jogosult intézménnyel, akkor ennek analógiájára mentális állapotaink mibenléte sem merülhet ki abban a tényben, hogy ezek az állapotok meghatározott tulajdonságokkal rendelkeznek a fejünkben, hanem az oksági leszármazás is meghatározza az identitásukat, az a kauzalitás, amely keletkezésüket a világ egy külső tényéhez köti. Oszkár fejében a „színtelen, szagtalan folyadék” fogalma a földi vízhez (tehát a H₂O-hoz) fűződő érzékelési viszony során alakult ki, és nem az XYZ-vel való oksági kapcsolatban, ilyen anyaggal ugyanis Oszkár soha nem találkozott. Ez a tény különbözteti meg az iker-Oszkár elméjében található fogalomtól.

Minthogy mentális állapotaink érzékenyek azoknak az okoknak a különbségére, amelyek létrehozták őket, egy tartályban táplált és egy eleven emberi testben található agy érzetei is eltérő oksági történettel fognak rendelkezni. Putnam ezért nem fogadja el az általa javasolt gondolatkísérlet (az „agyak a tartályban”) internalista következményeit. Még ha a szóban forgó érzetek belülről – internálisan – ugyanolyannak *tűnnek* is mindkét esetben, valójában nem azonos fajta mentális állapotok. A szuperkomputerrel indukált érzetek Putnam szerint éppúgy képtelenek arra a valóságra utalni, amelyhez okságilag semmi közük, mint ahogyan a földi Oszkár vízfogalma sem tud az XYZ-re referálni. *A mentális állapotokhoz tartozó utalásszerkezet (referencia) ugyanis olyan oksági viszony függvénye, amely az elmét a környezetéhez kapcsolja.*

Az externalizmus jelentősége

Az externalizmus az elmúlt évtizedekben új pályára állította az elmével kapcsolatos gondolkodást. Nem mintha a felsorolt érvek minden filozófust meggyőztek volna – az internalisták és az externalisták vitája ma is folyik.²⁰ Az az alapgondolat azonban, amely megfontolásra érdemes, a következő. Az elme nem olyan módon van jelen a testben, hogy a testi környezete pusztán valamilyen speciális, biológiai jellegű „tartály” lenne, valamilyen külső hordozó a tulajdonképpeni mentális folyamatokhoz képest, hanem a test és az elme bensőségesen összefonódik egymással. A test nem csupán okságilag hat az elme tartalmaira, hanem a testi, környezeti hatások konstitutív módon tartoznak az elme természetéhez. A *konstitutív* kifejezés arra utal, hogy egy dolog nem pusztán külső okként járul egy hatáshoz, hanem a hatás részévé válik. Okságilag a külvilág tárgyai az internalista szerint is hozzájárulnak az egyes mentális állapotok létrejöttéhez, de ez a tény – ha az adott állapot egyszer már létrejött – nem érinti annak mibenlétét. Amikor ezzel szemben az externalista azt mondja, hogy az elme oly módon része a bio- és ökoszisztémának, hogy ez a tény konstitutív az elme természetére nézve, ezen azt érti, hogy e tény nélkül az elme nem működhetne úgy, ahogyan működik, és nem lehetne *az, ami*.

Bár a kijelentés, hogy „*az elme a környező bio- és ökoszisztéma része*”, ugyanúgy hangzik, mint amikor a fizikalista állította, hogy az elme a fizikai világhoz tartozik, valójában másról van szó. Az externalizmus felől megfogalmazott állítás nem metafizikai kijelentés, nem a lélek és a test mibenlétével kapcsolatos (s ennyiben nem is ellentétes a fizikalista tézissel), hanem azt a *gyakorlati kapcsolatot* veszi célba, amely az elmét a környezetéhez köti.

²⁰ Az externalista érvek kritikájának világos összefoglalását lásd ADAMS & AIZAWA in ROBBINS & AYDEDE 2009, pp. 78-95. E vitákról kiváló és naprakész áttekintést ad GALLAGHER 2017, 1. fej.

2. A szituált kogníció

A klasszikus kognitivisták modellek internalisták: a mentális állapotokat a világról leválasztva, „fejben” lokalizálják, s így próbálják megérteni. Ezzel szemben az externalista úgy véli, hogy a mentális állapotok nem a „fejünkben vannak”. Egy korábbiakban megismert fogalomra utalva kissé elnagyoltan azt is mondhatjuk, hogy az externalizmus egy speciális ráépülési (szupervenienencia) tézis: a mentális jelenségek *bázisa egyszerre az organizmus és környezete*. Eszerint az elme nem független attól a testtől és attól a konkrét környezettől („szituációtól”), amelyben él és tevékenykedik, s ennyiben lényegénél fogva szituált megismerés jellemzi. E gondolat következményeit egy sor részletesebb elmélet és kutatási irány bontotta ki az utóbbi évtizedekben, melyek az alábbi négy angol hívószó (az ún. „4E”) köré csoportosíthatók: *embodied*, *embedded*, *enactive* és *extended*. E jelzők arra utalnak, hogy az elme *megtestesült*, *környezetébe ágyazott*, *enaktív* és *kiterjesztett* elmeként létezik a világban. Vegyük sorra e 4E-t és az egyes kifejezésekhez tartozó elméleteket.

Megtestesült és beágyazott elme

Milyen szerepet játszik a test a megismerésben? A klasszikus kognitívizmus modellje szerint a test funkciója kettős: egyrészt bemeneti információt (*input*) biztosít a mentális folyamatok számára, másrészt végrehajtja a kimeneti viselkedést (*output*). A test tehát az a közeg, amelyen keresztül kommunikálunk a környezetünkkel. Mint ilyen elkülönül a tulajdonképpeni kogníciótól, amely a be- és kimeneti tényezők közötti szférában zajlik („szendvicsmodell”). A test tehát – számítógépes nyelven fogalmazva – mintegy *perifériaként* van jelen.²¹ Az MI-kutatás ugyanakkor az utóbbi évtizedekben megmutatta, hogy e felfogás sok szempontból elhibázott. E modell alapján igen nehéz jól működő mesterséges kognitív rendszereket létrehozni, jellemző módon csak nagy energiafelhasználású és nehézkesen mozgó robotokat lehet az elv alapján előállítani. Ilyen humanoid volt a 2000-es évek elején a Honda cég Asimo robotja, melynek mozgása darabos, a járáshoz szükséges energiafelhasználása pedig (amellyel egységnyi tömeget egységnyi távolságra képes eljuttatni) rendkívül magas volt (lásd CLARK 2008, 1. fej.). Az ezzel kapcsolatos mérnöki problémákat végül nem a központi vezérlés kapacitásának növelése, hanem egy új szemlélet meghonosodása orvosolta: a vezérlés bizonyos feladatait egy önszabályozó, önmagát stabilizáló fizikai rendszer vette át. Ennek példája az úgynevezett Toddler robot, amely testi adottságai révén tud a környezetében navigálni (vö. uo.). Az elme csupán azt a teljesítményt segíti mint „facilitátor”, amit a test a maga morfológiai adottságainál fogva lehetővé tesz.

Milyen következtetés vonható le a fenti példákból? A Toddler robot esete arra mutat rá, hogy a környezetében tevékenykedő elmének sem kell minden mozgást „fejben” megterveznie, ezek egy részét nyugodtan rábízhatja a testére, hiszen a test a maga morfológiai és biológiai adottságainak köszönhetően képes ezeket megbízhatóan kivitelezni. A test ugyanis fizikai tulajdonságai révén közvetlen interakcióban áll a környezetével, amelyre mintegy önállóan ráhangolódik, s így az elme azokat a funkciókat, amelyekre a test saját dinamikus természetének köszönhetően képes, mintegy „kiszervezheti”. Ezeket a funkciókat az elmének nem kell költségesen, a saját kognitív erőforrásai révén működtetnie.

A testesültség elvének központi belátása, hogy a test és az elme között olyan munkamegosztás van, amely az elme működését mélyen befolyásolja (lásd VARELA; THOMPSON & ROSCH 1991). Elménk nem egy tetszőleges környezetben futtatható software, hanem a testtel közös evolúció során jött létre, így kölcsönösen ráhangolódtak egymás adottságaira. Elménk természete ily módon tükrözi a

²¹Emellett természetesen a kognitív folyamatok hordozójaként is szükség lehet rá, ez azonban – mint láttuk – a szimbólummanipulációként felfogott kogníció szempontjából nem releváns.

testet: olyan lélek, mint amivel mi emberek rendelkezünk, csak egy *ilyen és ilyen* – azaz emberi – testhez tartozhat.

Képzeljük el egy pillanatra, hogy ha az említett két robot érzékelné a testét, milyen tapasztalatokra tennének szert. Az Asimo saját testét ellenállást képező akadályként érzékelné, amelynek használata folyamatos, energiaigényes mikrovezérlést követel. Ezzel szemben a Toddler azt érezné, hogy elég akarnia, s a mozgás máris létrejön (CLARK 2008, p. 8). Ez lehetővé tenné számára, hogy optimális esetben a testtapasztalat mintegy háttérbe húzódjon, áttetszővé váljon, azaz a robot bizonyos körülmények között megfeledezhesen a testéről, pontosan úgy, ahogyan azt mi tapasztaljuk, amikor belefeledkezünk például egy izgalmas regénybe. A példa azt sugallja, hogy amikor intenzív szellemi tevékenység közben a test kikerül a figyelem fókuszából, ennek a tapasztalatnak az alapja nem a test lényegtelenítése a lelki jelenségekkel szemben, hanem épp ellenkezőleg, a testi folyamatok otthonossága. A test tudniillik ilyenkor is biztonsággal teszi a dolgát, otthon van a világban, s ezért nem szorul rá a mentális erőforrásokra. Ezt a viszonyt nevezi Andy Clark „otthonos interakciónak” (*inhabited interaction*), szemben a *leválasztott ellenőrzés* centralizált stratégiájával (*disconnected control*), amellyel a Honda-robot igyekezett megoldani a test mozgatásának komplex feladatát (vö. uo.)

Enaktív elme

A testesültség egy másik aspektusát az elme és a környezet viszonyának pragmatikus felfogása jelenti. Az *enaktivizmus* elmélete szerint ahhoz, hogy egy elmében kognitív folyamatok mehessenek végbe, a cselekvőnek aktív kapcsolatban kell állnia a környezetével (lásd NOË 2004, 1. fejj.; THOMPSON 2007). A környezet észlelése nem passzív befogadás, hanem cselekvés, s ott, ahol ennek a lehetősége megszakad, az észlelés is torzul. E megközelítés tehát azt hangsúlyozza, hogy elmével rendelkezni nem annyira állapot, mint inkább a világban való részvétel egy formája. Olyasfajta gyakorlat, mint a kerékpározás, ami jól begyakorolt jártasságot kíván, vagyis a „tudni, hogy hogyan” (a *know-how*) területére tartozik, és nem valamilyen elméleti tudás eredménye. Ez ismét más oldalról teszi kérdésessé az imént vázolt szendvicsmodellt, amelyben világosan elkülönült egymástól az *input–output*, valamint a köztes kogníció fázisa. Az enaktív elmélet e megkülönböztetés helyett a cselekvés és megismerés folyamatainak dinamikus összefonódását feltételezi.

Világítsuk meg ezt a látás példájával. A hagyományos észleléselemélet azoknak a perceptuális ítéleteknek a szerepét hangsúlyozza, amelyeknek köszönhetően a vizuális információkat értelmezzük. Értelmező ítéletekre eszerint azért van szükség, mert a retinán képződő ingermintázat kétdimenziós, azaz elszegényíti a valóságot, s így az optikai információból az agynak ugyanúgy kell „kiokoskodnia”, hogy a kétdimenziós kép milyen háromdimenziós testek vetülete, mint ahogy egy perspektivikus ábrát nézve is különböző vizuális jelzések alapján kell megítélnünk az ábrázolt tárgyak méretét és távolságát. Ezzel szemben az enaktív elmélet nagy súlyt fektet a szemlélő mozgására. Tipikus esetben nem állóképek értelmezésével alkotunk képet a világról, hanem folyamatosan mozgunk a környezetünkben, azaz szakadatlanul változtatjuk a nézőpontunkat, így a kétértelműségek kiküszöbölése egy heurisztikus gyakorlat révén történik. (Ha nem világos, hogy a látvány elemei milyen távolságra vannak egymástól, a legegyszerűbb a fejünket megmozdítani, s új szögből pillantani a tárgyra.) E tevékenység a legkevésbé sem rokon azzal, ahogyan egy falra szegezett perspektivikus képet igyekszünk megérteni. Az észlelő a megfigyelés aktív, mozgásos folyamatban legtöbbször mintegy ráhangolódik a vizuális információ áramára (vö. Gibson 1950, 1979). Aki például egy messzire repített labda után fut, az nem ballisztikai számítások alapján céloz meg egy pontot a pályán, hanem folyamatosan korrigálja mozgását a labdához képest, oly módon, hogy az információáram bizonyos típusú állandóságot mutasson.

Az enaktivizmus szerint az észlelés és a mozgás folyamatai összefüggő egységet alkotnak. Mindez természetesen tanulás eredményeként válik lehetségessé. Ám a tanulás során nem a projektív geometria vagy a ballisztika összefüggéseit sajátítjuk el, hanem szenzomotoros *skilleket* tanulunk. Azt tudniillik, hogy bizonyos típusú testi mozgások milyen típusú vizuális változásokat okoznak, s hogy ezek miként viszonyulnak a tárgyak várható téri viszonyaihoz. Az így elsajátított összefüggések az egyes érzéki modalitások (látás, hallás, tapintás...) szerint különböznek, s az adott modalitásra jellemző mintázattal rendelkeznek. A tapintási érzet például a sajáttest-észleléssel (a propriocepcióval) van szoros kapcsolatban: amikor az ujjunkat végighúzzuk egy tárgy felületén, érezzük a saját izmaink, végtagjaink változó állapotát. A látás ezzel szemben távolból történik. A látás során az egyes tárgyak takarásba kerülhetnek egymással, ám a takarás mértéke a saját test mozgásának függvényében változik, ami a hallásra vagy az ízelelésre nem jellemző. Sok jel mutat arra, hogy az egyes érzéki minőségek (vizuális, auditív, taktilis) jellege ezekhez a gyakorlatban elsajátított szenzomotoros mintázatokhoz tartozik, és nem magukhoz az érzékszervekhez (lásd NOË & HURLEY 2003).²² Ha a tapintás szerve olyan információkat kezd el továbbítani, ami a látásra jellemző mintázattal rendelkezik, az észlelő tapintás szervei révén vizuális információkra tesz szert. Paul Bach-y-Rita vakok hátára vagy nyelvére olyan elektródákat kapcsolt, amelyek egy fejre szerelt kamera jeleit fordították le különböző erősségű elektromos ingerekre az adott testfelületen (azaz a háton vagy a nyelven, lásd NOË 2009, pp. 56-57; CLARK 2008, 2.4. fej). A vakok arról számoltak be, hogy az ingerek egy rövid tanulási fázis után elvesztették taktilis jellegüket, mintegy „elváltak” a testtől, azaz a mozgás során érkező stimulusok térben távoli tárgyak vizuális képévé alakultak, s a kísérletben résztvevők képessé váltak távoli tárgyak felismerésére anélkül, hogy érintették volna őket. Fontos körülmény, hogy tanulás a kamera önálló mozgását igényelte. Ha más mozgatta a kamerát, az eredmény nem jött létre. Mindez ismét arra utal, hogy az észlelés nem passzív információfelvétel és ezt követő belső feldolgozás, hanem a világban végrehajtott aktív cselekvés. Az elme enaktív elmélete szerint a testben – a testnek köszönhetően – végzett motoros aktivitás nem választható el a szenzoros tevékenységtől. Az észlelés folyamatában ugyanis stabil visszacsatolási (feedback) körök jönnek létre: a mozgás manipulálása megváltoztatja az információforrást, az információáram megváltozása pedig visszahat a mozgásra. Az észlelés során a környezet és a megfigyelő csatolt rendszert alkot (lásd NOË 2004).

Kiterjesztett elme

A test–elme viszony újragondolásának legradikálisabb kísérletét az úgynevezett „kiterjesztett elme elmélete” jelenti a mai elmefilozófiában. Az ezzel kapcsolatos első gondolkísérlet David Chalmers és Andy Clark egy közös közleményében látott napvilágot 1998-ban. Inga és Otto egyaránt a New York-i Metropolitan Múzeumba tartanak. Inga jól emlékszik a címre és az útvonalra, Otto azonban előrehaladott Alzheimer-kórral küzd, ezért egy noteszt (*notebookot* vagy *tabletet*) használ erre a célra, amelyet mindig magánál hord, s amelyet *teljesen automatikussá* vált rutinnal kezel (lásd CHALMERS & CLARK in CHALMERS 1998/2002, pp. 643-652). A szerzők szerint nem adható az elme mibenlétének olyan jól indokolt definíciója, amely kizárná, hogy Otto noteszhasználatát pontosan ugyanolyan kognitív folyamatnak tekintsük, mint Inga memóriahasználatát. Az a különbség ugyanis, hogy Inga esetében az információ tárolása a koponyán belül, valamilyen szerves anyagban történik (vagy ha tetszik, valamilyen immateriális közegben is akár), míg Otto esetében szervesetlen, külső tárhelyen, nem jelent *elvi* különbséget abból a szempontból, hogy mindketten nehézség nélkül elő tudják hívni a

²² „Általánosságban fogalmazva: ami meghatározza a fenomenológiát, az nem az ingerlés által létrehozott idegi aktivitás mint olyan, hanem az, ahogy az idegi aktivitás betagozódik a szenzomotoros dinamizmusba.” (NOË 2009, p. 227)

megfelelő emlékeket, s ezeknek megfelelően hasonló sikerrel cselekszenek. Kognitív szempontból a két eset ekvivalens.

Ahhoz, hogy e gondolat kísérletet helyesen értsük, érdemes visszakanyarodni az enaktív elmélethez, amely megmutatta, hogy kognitív működéseink során állandó interakcióban állunk a környezetünkkel, s olyan csatolt rendszert alkotunk vele, amely lehetővé teszi az információ optimális áramlását. Itt is erről van szó. A kiterjesztett elme elmélete (továbbfejlesztett formában) nem azt sugallja, hogy egy notesz, egy tablet vagy egy könyv a szó szoros értelmében része egy személy elméjének, hanem azt, hogy a mentális folyamatok cselekvésként rögzíthetők, s ezért ráépülhetnek nemcsak az agy, hanem az agy, a test és a külső környezet egy megfelelő részére, mintegy összekapcsolva azokat. Amikor tehát cselekszünk, nincs értelme éles és merev határt húzni a megismerő szubjektum és az objektív világ között. Megismerni a dolgokat annyi, mint világgal összekötő *feedback* mechanizmusok révén cselekedni a környezetünkben. Innen nézve a gondolat kísérlet eredménye annak a felismerése, hogy mentális állapotaink nem statikus adottságok egy tudatnak vagy elmének nevezett lelki fakultásban, hanem aktív, dinamikus folyamatok.

Chalmers és Clark gondolat kísérlete egy olyan elmefogalom felé tett lépés, amelynek része a notesz és a felhasználó közötti *interakció* is, azaz az elme határai nem esnek egybe az individuum megszokott testhatáraival. Ez egy új, nem-karteziánus megismeréstudomány felé mutat, amely nem egyszerűen azt állítja, hogy a perceptuális folyamatok a környezetre „támaszkodva”, abba „beágyazva” folynak, hanem egy erősebb szupervenienencia-tézist fogalmaz meg: Bizonyos agyon kívüli tényezők a szó szoros értelmében konstitutív elemei a kognitív folyamatoknak, azaz nem csupán beágyazottságról, nem csupán a megismerést segítő okokról (testi „állványzatokról”) van szó, hanem arról, hogy bizonyos kognitív folyamatok (tehát nem az összes!) részben „ott kint” zajlanak a világban. Ez az ún. kiterjesztett elme elmélete (*extended mind theory*).

3. Az elmével kapcsolatos természetes kép és a szituált kogníció

A megtestesült, enaktív és kiterjesztett elme elméletei felől nézve sok tekintetben hamis az a kép, amelyet a mindennapi életben alakítunk ki magunknak az észlelésről. A természetes kép azt sugallja, hogy amikor kinyitjuk a szemünket, pontos és részletgazdag képet kapunk a környezetünkről. Számos kísérlet igazolja, hogy ez nem így van. Ismert tény például, hogy a látótér igen szűk területén rendelkezünk éles képpel, ott, ahol a fénysugarak a receptorokban gazdag foveára esnek a szemfenéken. Ennek ellenére az a természetes benyomásunk, hogy a vizuális világ *teljes terjedelmében* éles kontúrokkal látjuk a dolgokat. Kevésbé ismert tény, hogy a színérzékelő receptorok eloszlása a retinán nem egyenletes, ezért a látótér peremén egyáltalán nem látunk színeket. Ennek ellenére nincs olyan érzésünk, hogy a látvány egy része fekete-fehér. Kísérletek sora támasztja alá továbbá, hogy figyelmünk szűk, feladat-specifikus korlátok között szolgáltat információkat, ezért gyakran vakok vagyunk környezetünk szembetűnő elemeire. Ezt a jelenséget *figyelmetlenségi vakságnak* nevezzük. Jóllehet egyes testek ott vannak a tekintetünk előtt, ha a figyelmünket másvalami köti le, a szó szoros értelmében láthatatlanok számunkra (lásd CHABRIS & SIMONS 2011).

Gondoljunk bele: ha tudatunk valóban részletgazdag képet alakítana ki a környezetről – ahogyan azt például egy kamera teszi –, miként kerülhetné el a figyelmünket egy-egy szembetűnő esemény? Ha azonban az észlelés során nem részletgazdag belső modelleket alakítunk ki a világról, akkor vajon miért rendelkezünk mégis ezzel ellentétes hétköznapi meggyőződéssel? Miért van az a benyomásunk, hogy a látótér valamennyi pontja színes és pontos látványként adódik? A válasz az enaktív elmélet szerint abban keresendő, hogy a megismerés során a *cselekvési lehetőségek* horizontja helyettesíti a reprezentációkat. Igaz, hogy a kép csak egy kis folton éles a látótérben, de szemeink szakkadikus mozgása, valamint az állandó fej- és testmozgás a nézőpont szabad megválasztását teszi lehetővé, így

bármely részlet látványa bármikor a rendelkezésünkre áll éles és színes formában.²³ Ez a cselekvési szabadság kelti bennünk azt a benyomást, hogy egy részletesen észlelt környezetben mozgunk.

A szem szakkadikus mozgása és a szabad testmozgás révén nem egyszerűen befogadjuk a kínálkozó látványt, hanem aktív módon vagyunk jelen a világban. A környezet és a saját test manipulálásával mintegy kiprovokáljuk a szükséges észleleteket, és „letapogatjuk” a környezetünket. Ahogyan a vak ember a botjával tapogatja ki a környezetét, úgy mi is heurisztikus, exploratív tevékenységet végzünk az érzékeink segítségével (vö. NOË 2004, p. 95). A látás tehát cselekvés: a világban való jelenlétünk aktív formája.

III. Következtetések

Írásunk első részében az elmével kapcsolatos metafizikai vitákat tekintettük át, és arra a következtetésre jutottunk, hogy a fizikalista elméletek – különösen nem reduktív formájukban – nem cáfolhatók a magyarázati rés jelensége alapján. Nem mintha a fizikalista minden nehézségre válaszolni tudna, de az alternatív magyarázatok sem mentesek a nehézségektől. Úgy tűnik tehát, hogy negatív eredményre jutottunk. A második részében azonban eltérő irányból közelítettünk az elme és a természeti környezet viszonyához. Annak megmutatására törekedtünk, hogy a lélek mibenlététől függetlenül gondolkodásunk oly bensőségesen fonódik össze a környezeti folyamatokkal, hogy ez utóbbiak bizonyos körülmények között az elme konstitutív részének tekinthetők. Elmével vagy lélekkel rendelkezni tehát nem valamilyen sajátos metafizikai entitásként való létezést jelent, hanem dinamikus cselekvési formát, így az elme egy a környezetre nyitott rendszerként érthető meg. Milyen következményekkel jár mindez a bevezető sorokban feltett kérdések szempontjából? A teológiai-filozófiai hagyomány leggyakrabban internalista nézőpontból tekint a lélekre. Vajon miképpen módosulnak a kérdések és a lehetséges válaszok, ha az imént vázolt 4E felől pillantunk a jól ismert problémákra? A bemutatott szempontok alapján megkockáztatható, hogy a metafizikai interpretáció nehézségei (amelyek miatt negatív eredménnyel zártuk az első megközelítést) nem feltétlenül jelentenek valódi negatívumot. Az így hátramaradt bizonytalanság ugyanis természetes sajátossága azoknak a nyitott rendszereknek, amelyek nem írhatók le önmagukban zárt egységekként, azaz nem ragadhatók meg a különböző tárgyra jellemző ontológiai fogalmakkal. Edgar Morin szerint a bizonytalanság e fajtája termékeny ösztönzést jelent.

„Az ontológiai tévedés abban állt, hogy lezártuk, azaz megmerevítettük a tudomány (és a filozófia) alapvető fogalmait. Épp ellenkezőleg, olyan megismerés lehetőségét kell megnyitni, amely egyszerre gazdagabb és bizonytalanabb.” (MORIN 1990/2005, p. 60)

Mennyiben tekinthető azonban e bizonytalanság gazdagságnak?

Mint láttuk, a szituatív megismerés elsősorban azokra a fejlődéstanilag (és evolúciósan is) korai képességekre koncentrál, amelyek a környezettel való közvetlen interakciókat teszik lehetővé. A világban való aktív jelenlét gyakorlata, a szenzomotoros *skillek* elsajátítása és alkalmazása megelőzi azokat az elméleti ismereteket, amelyek teoretikusan vonatkoztathatók ugyanerre a világra. E körülmény lehetetlenné teszi, hogy az elmét izoláltan, azoktól a folyamatoktól elválasztva értsük meg, amelyek összekapcsolják a környezetével. Ez nem azt jelenti, hogy a léleknek ne lenne valamilyen autonómiája vagy belső világa, hanem azt mutatja, hogy az ágens épp azért képes kialakítani saját belső szféráját, mert konstitutív értelemben függ attól a környezettől, amelytől meg is különbözteti

²³E megközelítés szerint a környezet e folyamatos rendelkezésre állása feleslegessé teszi mindenre kiterjedő reprezentációk belső tárolását (vö. ROWLANDS 2010, p. 33).

magát. A belső a külsőtől való elhatárolás gyakorlatait követeli, így a mentális tartalmak valóban leírhatók „belsőként”, de csak egy előzetes megkülönböztetés függvényében.

„Miként egy folyó forgói, örvényei, turbulens struktúrái sem léteznek a sodrástól függetlenül, ugyanígy van ez az agy esetében is. A mentális létezők, szimbólumok, és más hasonlóak nem programozható tárgyként léteznek [...], hanem az agy soha nem szűnő dinamikus aktivitása teremti őket. A hiba, amit sok kognitív tudós elkövetett, abban áll, hogy a szimbolikus tartalmakat statikus, időtlen létezőknek tekintik, amelyek függetlenek az eredetüktől. A szimbólumok, miként a folyó örvényei, lehetnek szilárd és hosszan tartó struktúrák vagy mintázatok, de nem időtlenek és nem változatlanok.” (KELSO 1995, p. 1)

Az elme egysége és belső világa tehát kétségbevonhatatlan realitás, de éppúgy a környezet szakadatlanul változó folyamataitól függ, miként egy örvény azonossága és belső realitása is függ a víz folyamatos áramlásától. Ez a kettős viszony – függés és elhatárol(ód)ás – nemcsak az olyan bonyolult hálózatok sajátja, mint az agy, de valamilyen kezdetleges szinten minden élő szervezetre is jellemző, hiszen egy organizmusnak már a legegyszerűbb formában is különbséget kell tennie *külső* és *belső* között. Saját belső struktúráit oly módon kell fenntartania az állandóan változó környezetben, hogy bizonyos alapvető fizikai-biológiai értékek szűk határok között maradjanak (homeosztázis). Eközben a külvilág egyszerre veszély- és energiaforrás a számára: olyan állandóan változó hatások kezdőpontja, amelyhez az élőlénynek szünet nélkül alkalmazkodnia kell, de amelyekre struktúrái fenntartása és reprodukálása végett rá is szorul. A kogníció enaktív elmélete felől nézve az elme és a környezet viszonya is ilyen: a kívülről érkező stimulus nem csupán információ a világról, hanem a belső világ (a *milieu intérieur*) egyensúlyát megváltoztató hatás, perturbancia, amelyre az élőlénynek reagálnia kell.

Mindezek fényében meglepő lenne, ha a testi világban működő elme nem örökölné meg ezeknek a nyitott, dinamikus, önszabályozó rendszereknek az alapvető jellemzőit. Amikor a hagyományos filozófia a lélek metafizikai természetéről gondolkodik, olyan leírásokra törekszik, amely saját tulajdonságai felől próbálja megérteni a lelket. Az elmondottak azonban arra utalnak, hogy a *saját* és a *nem saját* közötti különbséget nem metafizikai adottságokként, hanem egy folyamatosan fenntartott, megkülönböztető gyakorlat eredményeként érdemes elgondolni. Az emberi elme nem külsődlegesen függ a fizikai, biológiai és társadalmi környezetétől, hanem oly módon fonódik össze velük, hogy a *saját* és a *nem saját* közötti határ dinamikusan változó tényezőket kapcsol össze.

Az így kapott lélekfogalom transzcendens: „több, mint önmaga”. De nem abban a metafizikai értelemben, hogy a lélek meghaladja a természeti képződményeket, hanem abban a speciális értelemben, hogy – miként az örvény az iménti példában – önálló valóságként sem zárulhat önmagára. Minthogy belső folyamatai a külsővel kontaktusban képződnek, az, ami rajta kívül van – az *idegen* –, bizonyos értelemben identitásának szerves része. Nem úgy, hogy abszorbeálja és önmagával azonossá teszi azt, ami más, hanem abban a korábban tisztázott jelentésben, hogy működéseit csak a környezetben lévő más(ik) által, vele kapcsolatban – vagy éppen vele oppozícióban – képes gyakorolni. Az *azonos* ezekben a rendszerekben mindig a *másra* utalt, így az elme természetének lehatárolása, önmagában vett meghatározása reménytelennek tűnik.

„Logikailag a rendszer csak akkor érthető meg, ha belefoglaljuk a környezetét, amely [a rendszer] számára egyszerre bensőséges és idegen, s oly módon része a rendszernek, hogy [eközben] hozzá képest külső.” (MORIN 1990/2005, p. 32)

Az elme e – mondjuk így – természetes transzcendenciája abban áll, hogy az „én”-ként vagy elmeként körülrajzolt belső világ a társadalomban adott *másikra*, a test biológiai környezetére, az ökoszisztémában adott erőforrásokra támaszkodik: ezek nélkül nem körvonalazható a természete. Annak megállapítása azonban, hogy a *másnak*, amelyre azonosságunk támaszkodik, hol vannak a

határai, nehezen volna lehetséges. Semmi sem zárja ki, hogy a természeti folyamatokra hangolt emberi létezés e nyitottságát a lehető legradikálisabban értelmezzük, s az elme enaktív jellegét összhangba hozzuk az Istenre mint abszolút *másikra* nyitott vallásfogalommal. Ez az összhang azonban – elkerülendő a spekulatív jelleget – nem metafizikai természetű lesz. Látszólagos feszültségben az évezredek teológiai és filozófiai hagyománnyal, az itt javasolt megközelítés a másik különböző megnyilvánulásait egy közösségi praxis keretein belül ragadja meg. Mit jelent ez?

E pragmatikus megközelítés annak a következménye, hogy az imént tárgyalt „4E” a magas intellektuális teljesítmények – az absztrakt, racionális gondolkodás – helyett azokra az alapvető perceptuális viszonyokra, szenzomotoros *skillekre* fókuszál, amelyek az emberi megismerőképességet összekötik az állatvilággal. Épp az ellenkező irányban keresi tehát a kiutat, mint a hagyományos megközelítés: a létezéssel kapcsolatos alapvető felismerések helyett a más(ik) különféle formáira való ráhangolódás lehetőségeit keresi a testi létezésről elkezdve egyre táguló körökben. Bizonyos értelemben azt látjuk ugyanis (elegendő a környezeti erőforrásokkal való visszaélés kortárs problémáira gondolnunk), hogy létezésünk nem a világ elméleti megértése terén, hanem az alkalmazkodás testközeli gyakorlataiban mutat alapvető hiányosságokat. Miért ne helyezhetnénk előtérbe a vallásnak azokat az aspektusait, amelyek – bármiképpen álljanak is a metafizika alapkérdései – a világban való „benne-létünk” gyakorlati módozatait állítják előtérbe? Miért ne foghatnánk fel a vallási létezést a *hogyan* kérdéseként, vagyis egy olyan praxisra tett javaslatként, amely az emberi létezés különböző lehetőségeire igyekszik temptatív formában *ráhangolódni*, köztük a reményen túli végső lehetőségre is? Euripidész drámái számos esetben fejeződnek be egy olyan többféleképpen variálódó rituális formulával, amely a vallási tapasztalat e sajátos aspektusát teszi láthatóvá:

„Sok alakja van az isteninek,
meglepetést az isten sokat ad,
és mire várunk, az sose jó el,
míg a sosevártnak utat nyit a sors.”

(EURIPIDÉSZ: *Alkésztisz*, ford. Devecseri Gábor)

E ráhangolódás a „sosevárt”-ra sokkal inkább a „tudni, hogy hogyan”, mint a „tudni, hogy mi” jegyében mozog.²⁴ Nem mintha a kettő elválasztható lenne egymástól. E kijelentések nem a vallás – vagy akár a mindennapi élet – kognitív dimenzióinak a kiiktatását célozzák, hanem (az *embodiment* kifejezésre utalva) e dimenziók megtestesítését egy praktikus viszony formájában.

Egy olyan Istenbe vetett hit számára, „aki a holtakat megeleveníti, és azokat, amik nincsenek, előszólítja, mint meglevőket” (Róm 4,17), a lélek anyagi vagy anyagtalan jellegéhez nem fűződnek alapvető érdekek. Az azonban döntő kérdés, hogy ez a hit hogyan illeszkedik világunk egyre komplexebbnek tűnő dinamikus hálózataiba. Az illeszkedésnek ugyanis számos módja van a szilárd identitásformáktól elkezdve a zárt önismereten alapuló, rigid cselekvési stratégiákon át a nyitott – az önazonosságot az állandóan változó társadalmi és természeti környezethez viszonyítva temptatív módon kereső – magatartásformáig. Míg az előbbiek a saját identitás mibenlétének tisztázására és megőrzésére koncentrálnak igyekeznek rögzíteni a vallásos egzisztencia határait, addig az utóbbiak a belsőt és a külsőt – az önazonost és az idegent – egymásra utalt rendszereknek tekintik. Ha az elmével

²⁴ A rituális gyakorlatokhoz mindig tartoznak elméleti magyarázatok, de megfontolandó Catherine Bell megállapítása, mely szerint a vallási közösségekben általában nagyobb – nemzedékeken átívelő – egyetértés van afelől, hogy bizonyos rítusokat *hogyan* kell elvégezni, mint arról, hogy az egyes rítuselemek *mit* jelentenek (vö. Bell 2009, pp. 182-187).

kapcsolatos iménti megfontolások megállják a helyüket, az utóbbi tűnik a transzcendencia iránti nyitottság minimális *természeti* feltételének.²⁵

IV. Irodalom – References

- ADAMS, F. & K. AIZAWA (2009): Why the Mind is still in the Head?. In: P. ROBBINS & M. AYDEDE (eds.): *The Cambridge Handbook for Situated Cognition*. Cambridge UP, Cambridge, pp. 78-95. doi:10.1093/mind/fzv212
- BAARS, B. J. (1988/1995): *A Cognitive Theory of Consciousness*. Cambridge UP, Cambridge.
- BELL, C. (2009): *Ritual Theory, Ritual Practice*. Oxford UP, Oxford, 2009.
- BLAKE, R. & R. SEKULER (2004): *Észlelés*. Osiris, Budapest. (ford. Csépe, V.)
- BLOCK, N. (1995/2002): Concepts of Consciousness. In: D. J. CHALMERS (ed.): *Philosophy of Mind. Classical and Contemporary Readings*. Oxford UP, Oxford, pp. 206-218.
- BLOCK, N. (1978): Troubles with Functionalism. *Minnesota Studies in The Philosophy of Science*, **9**: 261-325.
- BLOCK, N. (2004): What is Functionalism?. In: J. HEIL (ed.): *Philosophy of Mind. A Guide and Anthology*. Oxford UP, Oxford.
- CHABRIS, C. & D. SIMONS (2011): *A láthatatlan gorilla avagy hogyan csapnak be minket érzékeink*. Magnólia, Budapest. (ford. Hodász, E.)
- CHALMERS, D. J. (2005): A tudatos tapasztalat rejtélye. *Kellék*, **27-28**: 7-18. (ford. Telegdi-Csetri, Á.)
- CHALMERS, D. J. (2004): Szemközt a tudat problémájával. *Vulgo*, **5**(2): 14-35. (ford. Suttyák, T.)
- CHALMERS, D. J. (1996): *The Conscious Mind. In Search of a Fundamental Theory*. Oxford UP, Oxford.
- CHALMERS, D. J. & A. CLARK (1998/2002): The Extended Mind. In: D. J. CHALMERS (ed.): *Philosophy of Mind. Classical and Contemporary Readings*, Oxford UP, Oxford, pp. 643-652.
- DESCARTES, R. (1994): *Elmélkedések az első filozófiáról*. Atlantisz, Budapest. (ford. Boros, G.)
- GALLAGHER, S. & D. ZAHAVI (2008): *A fenomenológiai elme. Bevezetés az elmefilozófia és kognitív tudományba*. Lélekben Otthon Kiadó, Budapest. (ford. Váradi, P. & Török, T.)
- GALLAGHER, S. (2017): *Enactivist Interventions. Rethinking the Mind*. Oxford UP, Oxford. doi: 10.1093/oso/9780198794325.001.0001
- GIBSON, J. J. (1950): *The perception of the visual world*. Houghton Mifflin, Boston.
- GIBSON, J. J. (1979): *The ecological approach to visual perception*. Houghton Mifflin, Boston. doi: 10.4324/9781315740218
- HEIL, J. (ed.) (2004): *Philosophy of Mind. A Guide and Anthology*. Oxford UP, Oxford.
- HUSSERL, E. (2000): *Karteziánus meditációk*. Atlantisz, Budapest. (ford. Mezei, B.)
- KELSO, J. A. S. (1995): *Dynamic Patterns. The Self-Organization of Brain and Behavior*. The MIT Press, Cambridge (MA).
- KRIEGEL, U. (2007): Philosophical Theories of Consciousness. Contemporary Western Perspectives. In: P. D. ZELAZO; M. MOSCOVITCH & E. THOMPSON (eds.): *The Cambridge Handbook of Consciousness*. Cambridge UP, Cambridge, pp. 35-66. doi:10.1017/cbo9780511816789.004

²⁵ A tanulmány megírását az OTKA K 104574-es, 112542-es és 116234-es, valamint 120375-ös, 125012-es és 123839-es számú kutatási projektjei támogatták.

- LASHLEY, K. S. (1970): A tudat behaviorista értelmezése. In: KARDOS, L. (szerk.): *Behaviorizmus*, Gondolat, Budapest, pp. 59-82. (ford. Lénárt, E.)
- MCGINN, C. (1989/2004): Can We Solve the Mind-Body Problem?. In: J. HEIL (ed.): *Philosophy of Mind. A Guide and Anthology*. Oxford UP, Oxford, pp. 781-797.
- MORIN, E. (1990/2005): *Introduction à la pensée complexe*. Seuil, Paris.
- NAGEL, T. (2004): Milyen lehet denevérnek lenni. *Vulgo*, **5**(2): 3-12. (ford. Sutyák, T.)
- NOË, A. (2004): *Action in Perception*. The MIT Press, Cambridge (MA).
- NOË, A. (2009): *Out of Our Heads. Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. Hill and Wang, New York.
- NOË, A. & S. HURLEY (2003): Neural Plasticity and Consciousness. *Biology and Philosophy*, **18**: 131-168. doi:10.1023/a:1023308401356
- PLÉH, Cs. (2010): *A lélektan története*, Osiris, Budapest.
- PUTNAM, H. (1967): Psychological Predicates. In: W. H. CAPITAN & D. D. MERRILL (eds.): *Art, Mind, and Religion*. University of Pittsburgh Press, Pittsburgh, pp. 37-48.
- PUTNAM, H. (1975/2003): The Meaning of 'Meaning'. In: uő: *Mind, Language, and Reality*. Philosophical Papers 2., Cambridge UP, Cambridge, pp. 215-271. Magyarul megjelent (2010): A „jelentés” jelentése. *Különbség*, **10**(1): 13-74. (ford. Kovács, J. & Polgárdi, Á.)
- PUTNAM, H. (1981): Brains in a Vat. In: uő: *Reason, Truth, and History*. Cambridge UP, Cambridge, 1-21. Magyarul megjelent (2001): Agyak a tartályban. *Magyar Filozófiai Szemle*, **45**(1-2): 1-22. (ford. Ruzsa, F.) doi:10.1017/cbo9780511625398.003
- ROWLANDS, M. (2010): *The New Science of the Mind. From Extended Mind to Embodied Phenomenology*. The MIT Press, Cambridge (MA). doi:10.7551/mitpress/9780262014557.001.0001
- SEARLE, J. R. (1980): Minds, Brains, and Programs. *Behavioral and Brain Sciences*, **3**: 417-424. doi:10.1017/s0140525x00005756
- SEARLE, J. R. (2007): *Freedom and Neurobiology*. Columbia UP, New York.
- SEARLE, J. R. (2008): *Philosophy in a New Century. Selected Essays*. Cambridge UP, Cambridge. doi:10.1017/cbo9780511812859
- THOMPSON, E. (2007): *Mind in Life. Biology, Phenomenology, and the Sciences of Mind*. The Belknap Press of Harvard UP, Cambridge (MA).
- TURING, A. M. (1950): Computing Machinery and Intelligence. *Mind*, **49**: 433. doi:10.1093/mind/lix.236.433
- VARELA, F. J.; E. THOMPSON & E. ROSCH (1991): *The Embodied Mind. Cognitive Science and the Human Experience*. The MIT Press, Cambridge (MA).
- WATSON, J. B. (1970): Hogyan látja egy behaviorista a pszichológiát?. In: KARDOS, L. (szerk.): *Behaviorizmus*. Budapest, Gondolat, pp. 39-58. (ford. Kulcsár, Zs.)
- WITTGENSTEIN, L. (1998): *Filozófiai vizsgálódások*. Atlantisz, Budapest. (ford. Neumer, K.)

